DOCUMENT RESUME

ED 081 295 FL 004 757

TITLE	Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, 1 April - 30 June 1973.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Dental Research (NIH), Bethesda, Md.; Office of Naval Research, Washington, D.C. Information Systems Research.
REPORT NO	SR-34-73
PUB DATE	Jun 73
NOTE	219p.
EDRS PRICE	MF-$0.65 HC-$9.87
DESCRIPTORS	Articulation (Speech); Audition (Physiology); Child Language; Cognitive Processes; Feedback; Information Processing; Intonation; Language Learning Levels; Language Patterns; *Language Research; Memory; Perception; Phonetics; *Phonology; *Physiology; *Research Tools; Socioeconomic Influences; *Speech; Stimulus Devices; Vowels

ABSTRACT

This document, containing 15 articles and 2 abstracts, is a report on the current status and progress of speech research. The following topics are investigated: phonological fusion, phonetic prerequisites for first-language learning, auditory and phonetic levels of processing, auditory short-term memory in vowel perception, hemispheric specialization for speech perception, oral feedback under anesthesia, laryngeal control in Korean stop production, and intonation in speech. A list of related publications and reports is also included. (DD)

FILMED FROM BEST AVAILABLE COPY

ED 081295

SR-34 (1973)

SPEECH RESEARCH

A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Applications

1 April - 30 June 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

FL004 757

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

Information Systems Branch, Office of Naval Research
Contract N00014-67-A-0129-0001 and -0002

National Institute of Dental Research
Grant DE-01774

National Institute of Child Health and Human Development
Grant HD-01994

Research and Development Division of the Prosthetic and Sensory Aids Service, Veterans Administration
Contract V101(134)P-71

National Science Foundation
Grant GS-28354

National Institutes of Health
General Research Support Grant RR-5596

National Institute of Child Health and Human Development
Contract NIH-71-2420

The Seeing Eye, Inc.
Equipment Grant

CONTENTS

I. Manuscripts and Extended Reports

Levels of Processing in Phonological Fusion -- James E. Cutting 1

Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation Modes -- James E. Cutting 55

Phonological Fusion of Stimuli Produced by Different Vocal Tracts -- James E. Cutting 61

Phonetic Prerequisites for First-Language Acquisition -- Ignatius G. Mattingly 65

A Note on the Relation between Action and Perception -- M. T. Turvey . . . 71

Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing -- David B. Pisoni and Jeffrey Tash 77

The Role of Auditory Short-Term Memory in Vowel Perception -- David B. Pisoni 89

Effects of Amplitude Variation on an Auditory Rivalry Task: Implications Concerning the Mechanism of Perceptual Asymmetries -- Susan Brady-Wood and Donald Shankweiler 119

Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects -- Ruth S. Day 127

On Learning "Secret Languages" -- Ruth S. Day 141

Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes -- M. F. Dorman and Donna S. Geffner 151

Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech -- Gloria Jones Borden, Katherine S. Harris, and William Oliver 159

Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia -- Gloria Jones Borden, Katherine S. Harris, and Lorne Catena 167

Laryngeal Control in Korean Stop Production -- Hajime Hirose, Charles Y. Lee, and Tatsujiro Ushijima 191

Patterns of Palatoglossus Activity and Their Implications for Speech Organization -- F. Bell-Berti and H. Hirose 203

ABSTRACT: Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency -- James E. Atkinson 211

ABSTRACT: Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses -- Charles C. Wood 215

II. Publications and Reports 219

III. Appendix: DDC and ERIC numbers (SR-21/22 - SR-33) 222

I. MANUSCRIPTS AND EXTENDED REPORTS

Levels of Processing in Phonological Fusion*

James Eric Cutting+

ABSTRACT

Phonological fusion occurs when the phonemes of two different speech stimuli are combined into a new percept which is longer and linguistically more complex than either of the two inputs. For example, when PAY is presented to one ear and LAY to the other, the subject often perceives PLAY. The purpose of the present studies was to determine whether both higher-level linguistic cues and lower-level nonlinguistic cues were responsible for fusion. Fusible stimuli were varied along linguistic and nonlinguistic dimensions in order to determine the level at which the information from the two inputs is combined into a single percept.

Fusion occurred independent of wide variations in the nonlinguistic dimensions of the stimuli. When to-be-fused stimuli were varied in their relative onsets by 100 msec or more, fusion still occurred at a high rate. Pitch differences of 20 Hz and intensity differences of 15 db had no effect on fusion rate. Insensitivity to these nonlinguistic stimulus dimensions is a characteristic of higher-level processes.

Although it was independent of nonlinguistic cues, phonological fusion was influenced by many kinds of linguistic cues. At the semantic level, fusible pairs yielded higher fusion rates when imbedded in a sentence context than when presented as isolated pairs. At the phoneme level, fusion rates were higher for certain phonemes than for others: for example, stop/liquid pairs such as BED/LED fused much more readily than fricative/liquid pairs such as FED/LED. While the particular phoneme chosen as the first consonant in a to-be-fused

---

*A dissertation presented to the faculty of the Graduate School of Yale University in candidacy for the degree of Doctor of Philosophy (Psychology), 1973.

+Haskins Laboratories and Yale University, New Haven, Conn.

Acknowledgment: I wish to thank Ruth S. Day, with whom I have never had a colorless conversation; Franklin S. Cooper, with whom I have never had a fruitless conversation; and the members of the New Haven Dance Ensemble, with whom I have never talked about this at all.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

cluster played an important role in fusion, the second played a less clear role. All liquid (/r,l/) and semivowel (/w,y/) stimuli fused equally well when paired with an appropriate stop consonant. At the acoustic level, specific cues were also important in facilitating fusion. For example, the second formant transition of the liquid stimulus (or one similar to it) was necessary, but not entirely sufficient for fusion to occur.

Thus, phonological fusion was insensitive to nonlinguistic stimulus variation, but sensitive to linguistic variation. These findings suggest that phonological fusion is a higher-level phenomenon. Moreover, they lend further support to the view that there are different processing mechanisms for linguistic and nonlinguistic dimensions.

INTRODUCTION

Most of the dichotic listening literature has dealt with the phenomenon of perceptual rivalry. Given a different stimulus presented to each ear at the same time, the subject typically reports hearing one or both of the stimuli. The different information contained in each stimulus is not combined into a single percept. Thus for example, given the dichotic digits ONE/FIVE, the subject never reports hearing FUN or WIVE. Perceptual fusion does occur when certain variables are taken in account. In several types of fusion phenomena the stimulus variables which facilitate fusion appear to be psycholinguistic in nature. For example, given the dichotic pair BANKET/LANKET the subject often reports hearing BLANKET (Day, 1970a). In this type of fusion segments of both stimuli are combined to form a new percept which is longer and linguistically more complex than either of the two inputs.

This phenomenon is called phonological fusion because it conforms to the phonological rules of English: given BANKET/LANKET the subject reports hearing BLANKET, not LBANKET. According to phonological rules of cluster formation in English, initial stop consonant + liquid clusters are allowed but initial liquid + stop clusters are not. Day (1970b) found that when stimuli were paired such that these constraints were removed, fusion occurred in both directions. Given the stimuli TASS/TACK, for example, the subject reported hearing TASK on some trials and TACKS on others. Both /sk/ and /ks/ clusters are permissible in final position in English.

Phonological fusion cannot be explained as a response bias for acceptable English words. Day (1968) found that when different productions of BANKET were presented to each ear, subjects reported hearing BANKET. That is, they did not report hearing the acceptable English word that corresponded most closely to the nonword inputs. Likewise, LANKET/LANKET yielded LANKET. Only when the stimuli were BANKET/LANKET did subjects report hearing BLANKET, regardless of which stimulus was presented to each ear.

Fusion also cannot be explained in terms of subjects' expectations, DAY (in preparation-a) has informed subjects before the fusion task about the type of stimuli they were to hear: among other items, some pairs consisted of different productions of BLACK (BLACK/BLACK), and some consisted of BACK and LACK (BACK/LACK). Given the opportunity to write down BLACK, BACK, or LACK as a response for any trial, subjects typically reported hearing BLACK in both conditions.

A Levels-of-Processing Approach to the Study of Cognition

There are two basic experimental approaches to the systematic study of process levels in cognition. In one approach the experimenter varies the task while holding the stimuli constant. In the second approach the experimenter varies the stimuli while holding the task constant. Typically, in the task-variation strategy the experimenter requires the subject to process different dimensions of the same stimuli. Day and Cutting (1970); Day, Cutting, and Copeland (1971); Wood, Goff, and Day (1971); Day and Wood (1972); and Wood (1973) have used such an approach. In all these studies stimuli were chosen so that in one task subjects were required to process linguistic dimensions of the stimuli, while in another task they were required to process nonlinguistic dimensions. In the stimulus-variation approach stimuli are varied along linguistic dimensions in some cases and nonlinguistic dimensions in others. The effects of the different types of variation are measured in the results of a common task. Day and Cutting (1971) and Cutting (1973c) have used this approach for tasks involving dichotic rivalry.

The present experiments used the stimulus-variation approach to study dichotic phonological fusion. Overall fusion level (or fusion rate) was the primary dependent variable. By varying the stimuli in a fusible pair along a particular dimension and observing fusion rate for varied and nonvaried fusible pairs, the level at which information is combined from the two inputs can be determined. Nonlinguistic variables, such as timing, pitch, and intensity, are considered first; linguistic variables, such as semantic context, the particular phonemes to be fused, and the acoustic structure of the stimuli, are examined second. In order to assess the importance of linguistic and nonlinguistic parameters stimuli must be carefully controlled. For example, in considering the importance of pitch on fusion rate, care must be taken so that there is no uncontrolled variation along another dimension, such as intensity or duration. Precise variation along linguistic and nonlinguistic dimensions is most readily achieved through the use of synthetic speech. Therefore, synthetic speech stimuli were used in all experiments. Although synthetic speech typically sounds somewhat artificial, especially to naive subjects, Cutting (1973a) found that the rules that governed the fusibility of synthetic stimuli and natural speech stimuli were the same, and that there were no inherent artifacts in the perception of synthetic speech which affected fusion results.

Types of Auditory Fusion

Since the 1950's a number of dichotic phenomena have been called fusion by various researchers. Broadbent (1955), Day (1968), and Halwes (1969), among others, have described experimental situations in which two auditory signals presented separately to each ear were perceived as one. From the titles of their papers one would assume that they were concerned with the same process: "On the fusion of sounds reaching different sense organs" (Broadbent and Ladefoged, 1957); "Fusion in dichotic listening" (Day, 1968); and "Effects of dichotic fusion on the perception of speech" (Halwes, 1969).

Fusion, however, is not one phenomenon, but many phenomena which are only tenuously related. Subsuming them all under the single label fusion with no descriptive adjective easily leads to confusion. Cutting (1972) has described six different types of auditory fusion and argued that they can be divided into two general groups according to their relative sensitivity to several stimulus

parameters. The groups are designated lower-level and higher-level fusions. Examples of the stimuli and possible fusion responses for each type are shown in Figure 1.

Lower-level fusions are dependent on nonlinguistic dimensions of the stimuli. If timing, in terms of the relative onset times of the two dichotic stimuli, is varied within a very small range fusion disintegrates and two stimuli are heard. For lower-level fusions this range is often a matter of microseconds. In addition, small differences between the stimuli in pitch (2 Hz) or intensity (2 db) are often sufficient to eliminate fusion so that the two stimuli are perceived as separate entities.

Higher-level fusions are relatively independent of these stimulus dimensions. Timing (relative onset time) differences of 25 msec are often insufficient to reduce fusion rates. Pitch and intensity may also vary between the two stimuli within a much greater range: differences of 20 Hz or 30 db may not reduce fusion rate at all. It appears that information in the stimuli--not relative onset time, pitch, or intensity--is important to fusion at higher levels.

Listed below are the six types of auditory fusion shown in Figure 1 along with a brief description of the phenomenon involved in each. Table 1 summarizes the relative sensitivities of five fusions to the nonlinguistic parameters of time, pitch, and intensity. For a more complete discussion of each fusion and for comparisons among them see Cutting (1972).

TABLE 1: Nonlinguistic dimensions which are relevant for the separation of lower-level and higher-level fusions. Tolerances of stimulus variation are listed within each cell. Specific numbers reflect current knowledge.

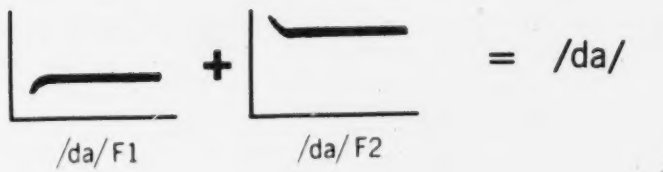
	Timing	Pitch	Intensity
LOWER-LEVEL FUSIONS			
1. Sound Localization	<2 msec	<2 Hz	<2 db
2. Spectral Fusion	<5 msec	<2 Hz	*
3. Psychoacoustic Fusion	*	<2 Hz	*
HIGHER-LEVEL FUSIONS			
4. Phonetic Feature Fusion	25 msec	>20 Hz	>30 db
5. Chirp Fusion	25 msec	>20 Hz	>30 db

*Systematic data not available.

1. Sound localization occurs for all audible sounds, speech and nonspeech. The first display of Figure 1 shows that when /da/ is presented to both ears at the same time, pitch, and intensity, the subject perceives one /da/ localized at

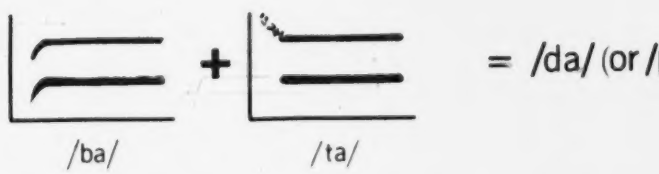
1. SOUND LOCALIZATION

2. SPECTRAL FUSION



3. PSYCHOACOUSTIC FUSION

4. PHONETIC FEATURE FUSION



5. CHIRP FUSION

6. PHONOLOGICAL FUSION

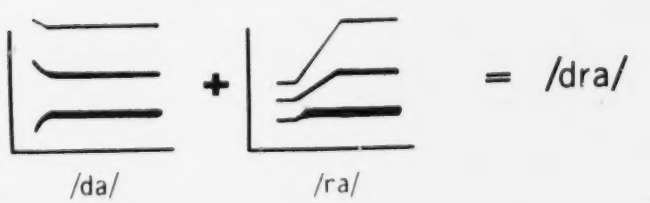


Figure 1: Six fusions of /da/. Schematic spectrograms of speech and speech-like stimuli in six types of auditory fusion.

the midline. Timing differences of 2 msec and pitch differences of 2 Hz are sufficient to cause the fused percept to disintegrate into two elements. Intensity differences of 2 db are sufficient to change the fused percept.

2. Spectral fusion occurs for speech sounds and for complex nonspeech sounds. For example, when the first formant (F1) of /da/ is presented to one ear and the second formant (F2) to the other, subjects perceive the fused /da/ as if it had undergone no special presentation technique. Timing differences of 5 msec and pitch differences of 2 Hz are sufficient to disrupt fusion so that two items are heard.

3. Psychoacoustic fusion probably occurs for both speech and nonspeech sounds, but only speech stimuli have been considered in an experimental situation. For example, when /ba/ is presented to one ear and /ga/ to the other, subjects often report hearing the fusion /da/. Such a fusion can only be accounted for by averaging the second formant (F2) transitions of /b/ and /g/. Pitch is the only dimension which has been explored experimentally: differences of 2 Hz are sufficient to inhibit fusion.

4. Phonetic feature fusion occurs only for competing speech segments. When /ba/ and /ta/, for example, are presented to opposite ears, subjects report hearing a "blend" of the phonetic features in the stimuli: /da/ or /pa/. Such responses involve extracting the voicing feature from one stop consonant and combining it with the place feature of the other stop. These responses cannot be accounted for by acoustic averaging. Timing differences of 25 msec do not disrupt this form of fusion, although greater differences decrease fusion rate. Pitch differences as large as 20 Hz and intensity differences of 30 db appear to have little effect on fusion rate.

5. Chirp fusion is demonstrated in the fifth display of Figure 1. If the second formant transition is separated from /da/, it sounds like a pitch sweep, rather similar to a bird's twitter; hence the name "chirp." The remainder of the stimulus, the "chirpless" /da/ sounds rather ambiguous, and resembles /ba/ more than /da/. When the chirp stimulus is presented to one ear and the "chirpless" /da/ to the other, the subject often reports hearing a complete /da/ plus a nonspeech chirp. Hence the chirp is perceived in two forms at the same time. Pilot work suggests that timing differences of 25 msec do not inhibit fusion, although greater relative onset times reduce fusion rate. Pitch differences of 20 Hz and intensity differences of 30 db have little or no effect on fusion rate.

6. Phonological fusion occurs for pairs of phonemes which can form permissible clusters; for example, when /da/ is presented to one ear and /ra/ to the other, the subject often perceives /dra/. Previous studies have shown that phonological fusion is tolerant to lead-time differences of as much as 150 msec (Day, in preparation-b). Experiment I was designed to test fusion at longer lead-time differences. The effects of pitch and intensity variations between the stimuli had not been measured. Experiments II and III were designed to obtain this information and to aid in the classification of phonological fusion as a higher- or lower-level fusion.

GENERAL METHODOLOGY

Terms

1. Stop stimulus. The member of a fusible pair which begins with a stop consonant, for example, PAY.

2. Liquid stimulus. The member of a fusible pair which begins with a liquid, for example, RAY and LAY.

3. Fusion response. A combination of phonemes from each ear into a fused cluster, for example, PLAY.

4. Lead time. The temporal interval between the onsets of the stimuli in a dichotic pair. Relative onset time and stimulus onset asynchrony are terms synonymous with lead time.

5. Dichotic presentation. The presentation of two different stimuli, one to each ear.

6. Diotic presentation. The presentation of the same stimulus to both ears at the same time such that the subject perceives one stimulus localized at the midline. Often this type of presentation has been called binaural; however binaural presentation is a general term denoting the stimulation of both ears at the same time. Thus, in a strict sense, both dichotic and diotic presentations are examples of binaural stimulation (Licklider, 1951:1027).

Conventions

Capital letters are used to indicate both stimuli and responses. For example, PAY and LAY often yielded the fused response PLAY. A slash between two stimuli indicates a dichotic pair (PAY/LAY), and an arrow (⟶) should be read as "yields;" thus, PAY/LAY⟶PLAY. Phonemes are indicated in lower case letters between a pair of slashes, such as /r/ and /l/.

Method

Stimuli and tapes. Synthetic stimuli used in all experiments were generated on the Haskins Laboratories' parallel resonance synthesizer. They were first synthesized by rule and then modified using the computer-controlled EXECUTIVE system (Mattingly, 1968) so that all parameters of the synthetic speech more nearly resembled those found in natural speech. Natural speech parameters were gathered from a number of sources including spectrograms, oscillograms, and published sources of parameter values (Lisker, 1957; O'Connor, Gerstman, Liberman, Delattre, and Cooper, 1957; Liberman, Ingemann, Lisker, Delattre, and Cooper, 1959; and Lehiste, 1962). Specific aspects of the acoustic structure of the stimuli are explained in Appendix A and in Cutting and Day (1972).

Synthetic stimuli were transferred to the pulse code modulation (PCM) system (Cooper and Mattingly, 1969) where they were digitized and stored on disc file for the preparation of experimental tapes. The PCM system allows the experimenter to record two stimuli simultaneously on the two channels of an audio tape, and to specify various relative onset values for the stimuli in each dichotic pair. Most dichotic tapes used three relative onset times: the stop

stimulus (e.g., PAY) began 50 msec before the liquid stimulus (e.g., LAY), the two stimuli began simultaneously, or the liquid stimulus began 50 msec before the stop stimulus.[1] The lead times occurred with equal probability and in a random order. The accuracy of relative onset times was within .5 msec. Channel arrangements for stimuli in fusible pairs were always counterbalanced within a dichotic tape; for example, on half the trials PAY was recorded on channel A and LAY was recorded on channel B, while on the other half of the trials the reverse configuration was recorded.

Diotic tapes were prepared for identification tasks in order to assess the extent to which each stimulus could be identified in isolation.

Subjects and apparatus. One hundred twelve Yale University undergraduates participated in nine experiments. Each received course credit for his or her services. The subjects were all right-handed native American English speakers with no history of hearing difficulty. They listened, generally in groups of four, to tapes played on an Ampex AG-500 dual track tape recorder. Auditory signals were sent through a listening station to Grason-Stadler earphones (Model TDH39-300Z). Gains on the tape recorder and listening station were adjusted so that stimuli were presented at approximately 70 db sound pressure level.[2] Earphone assignments were counterbalanced within subjects when possible, and across subjects when experimental tapes were too lengthy to permit the within-subject control.

Procedure. In most experiments, subjects participated in two tasks: a dichotic fusion task and a diotic identification task. In all cases the subjects' first task was the fusion task. The experimenter read them the following standard instructions for phonological fusion tasks as they read silently from their own copies:

> This is an experiment in speech perception. You will be listening to a series of messages through earphones. After each presentation, you are to write down what you heard. You will have to respond immediately for there will only be a few seconds before the next presentation begins.
>
> In order to do a good job, you must report exactly what you heard. For example, if you heard a real word, write down that real word; if you heard a nonsense word, write that nonsense word; if you heard one word, write it; if you heard two words, write them both; and so on. If you are not sure about what you heard, make a guess anyway: you must write something down after every presentation.
>
> Some of the items may sound very similar to others; however, they may in fact be different. Therefore, be careful to judge each presentation on its own merits.

---

[1] Experiment I used other lead times as well, and Experiment IV used only the simultaneous presentation.

[2] The only exception was Experiment III where intensity level was experimentally varied.

Before the task began subjects listened to several practice pairs and wrote their responses in order to familiarize themselves with the task and the stimuli. After listening to the practice pairs subjects were typically curious about the stimuli. They were reassured that the trials sometimes sounded odd to subjects, but that they should perform the task as best they could. If questions arose after listening to practice items subjects were referred back to the instruction sheet. They were not told that different stimuli were presented to different ears until after the completion of the fusion task.

In most experiments a second task was also run: diotic identification of individual stimuli. Single items such as PAY, RAY, and LAY were presented in a random sequence and subjects wrote down what they heard. The results of the identification tasks were consistent across all experiments: the individual stimuli were highly identifiable. For the sake of flow in the discussion of the fusion experiments, the results of the diotic identification tasks are summarized in Appendix B.[3] Identification tasks always followed fusion tasks so that subjects were not given precise knowledge about the stimuli before the fusion task began.

The statistical significance of a result was determined by a sign test on the individual subjects' scores. The z scores and p values are given only when results were significant at the .05 level or less.

I. NONLINGUISTIC DIMENSIONS

Experiment I: Timing

Timing appears to be an unimportant factor in phonological fusion. Using disyllabic natural speech stimuli Day and Cutting (1970) and Day (in preparation-b) found that phonological fusion was remarkably insensitive to differences in relative onset times of as much as 150 msec. Their results showed that fusion occurred almost as readily when the liquid stimulus (e.g., LANKET) led the stop stimulus (BANKET) by 150 msec, as when the stop stimulus led the liquid by the same extent. Furthermore, fusion rates for both cases were nearly identical with that for the simultaneous onset case. The longest relative onset time studied to date is 150 msec. The present study examined even longer lead times in order to determine the interval at which fusion rate drops substantially.

Method. Two fusion sets of the same general pattern were selected: the PAY set (PAY, RAY, LAY) and the KICK set (KICK, RICK, LICK). Members of the PAY set were 350 msec in duration, and members of the KICK set were 325 msec in duration. Dichotic pairs were assembled for all combinations of fusible items within a set: pairs and possible fusions were PAY/RAY⟶PRAY, PAY/LAY⟶PLAY, KICK/RICK⟶CRICK, and KICK/LICK⟶CLICK. Eleven lead times were selected: 0, $\pm$ 50, $\pm$ 100, $\pm$ 200, $\pm$ 400, $\pm$ 800. Plus and minus signs refer to relative onset times for the same stimuli: the plus refers to pairs in which the stop stimulus led the liquid, while the minus refers to liquid-leading pairs. There were equal numbers of stop-leading and liquid-leading trials. Since the longest stimuli (the PAY set) were only 350 msec in duration, the 400 and 800 msec conditions involved temporally non-overlapping stimuli for both sets. All fusible

---

[3] The results of the identification task in Experiment VII are discussed in the text. No identification tasks were run in Experiments IV and IX.

pairs were assembled into two independent tapes, each with a different random order. Each tape consisted of 88 pairs: (2 sets of stimuli) x (2 stop/liquid pairs per set) x (11 lead times) x (2 channel arrangements per pair). The order of listening to tapes was counterbalanced across eight subjects.

Major results. As shown in Figure 2, fusion occurred most readily when the stop stimulus led the liquid by 50 and 100 msec, where fusion rates were 63 and 59 percent, respectively. Fusion rate dropped substantially at the - 200, ± 400, and ± 800 msec leads, where fusion averaged 10 percent. Intermediate fusion rates were observed at the 0, - 50, - 100, and + 200 msec leads. No subject deviated markedly from the group data.

Fusion rates at the short leads (0, ± 50) were comparable to those found in previous studies using the same stimuli (Cutting, 1973a, b), and to those found in other experiments in this paper. Hence fusion rate was stable for short-lead items even when they appeared in a sequence with long-lead items that rarely fused. The fusions that did occur at long leads occurred primarily at the beginning of the task, and rapidly diminished thereafter. Fusion for shorter leads, however, continued at the same high rate throughout the task.

The formant transitions in the stop stimuli were about 50 msec in duration, while those in the liquid stimuli were 150 msec in duration. These segments did not need to overlap in time in order for fusion to occur, since fusion rate was substantial for the + 100 msec lead case.

Other results. In previous studies using the same stimuli at short leads of 0 and ± 50 msec, subjects usually reported a single item when they did not fuse; for example, PAY (Cutting, 1973a, b). The liquid stimulus was rarely reported in such cases. One-item responses, including fusions (PLAY) and nonfusions (PAY), accounted for more than 88 percent of all responses. In the present study, however, a wide variety of responses occurred: three kinds of one-item responses (PLAY, PAY, or occasionally LAY), and a large percentage of two-item responses (PAY and LAY, PLAY and PAY, PLAY and LAY, or occasionally PLAY and PLAY). One-item responses occurred predominantly at short leads while two-item responses occurred at the long leads. Figure 3 compares the one-item responses at each lead with the fusion responses shown in Figure 2. The one-item response curve included both fusions and nonfusions, and was generally symmetrical: an equal number of one-item responses occurred when the stop stimulus led as when the liquid stimulus led. The fusion response curve was not symmetrical, since considerably more fusions occurred when the stop stimulus began first.

The effects of relative onset time on phonological fusion may be summarized by two principles. First, fusions occurred frequently when the stimuli had relative onsets of ± 100 msec or less, but infrequently when the relative onsets were ± 200 or more. The mean value between these relative onset values is 150 msec, a lead time which may well be the maximum relative onset at which fusion will occur for these stimuli. High rates of fusion do occur at ± 150 msec for disyllabic natural speech stimuli (Day and Cutting, 1970; Day, in preparation-b). Second, fusions occurred more readily when the stop stimulus led the liquid than in the reverse configuration. The extent to which the first principle governs fusion rate is probably a function of the syllable structure and duration of the stimuli. In fact there may be a direct relationship between stimulus duration and the maximum relative onset time at which fusions occur. In

Figure 2

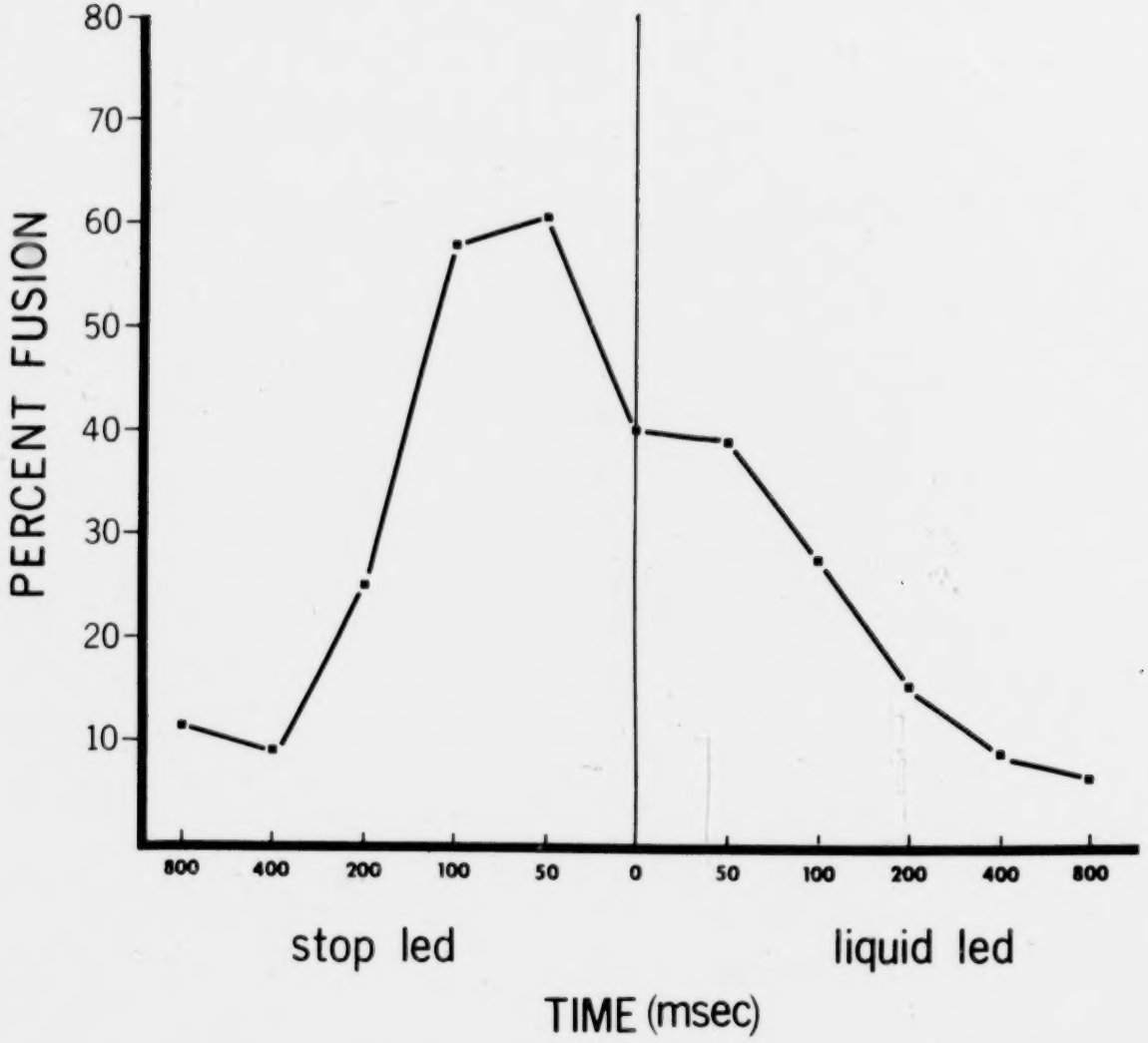


Figure 2: Overall fusion rate at eleven different lead times.

Figure 3

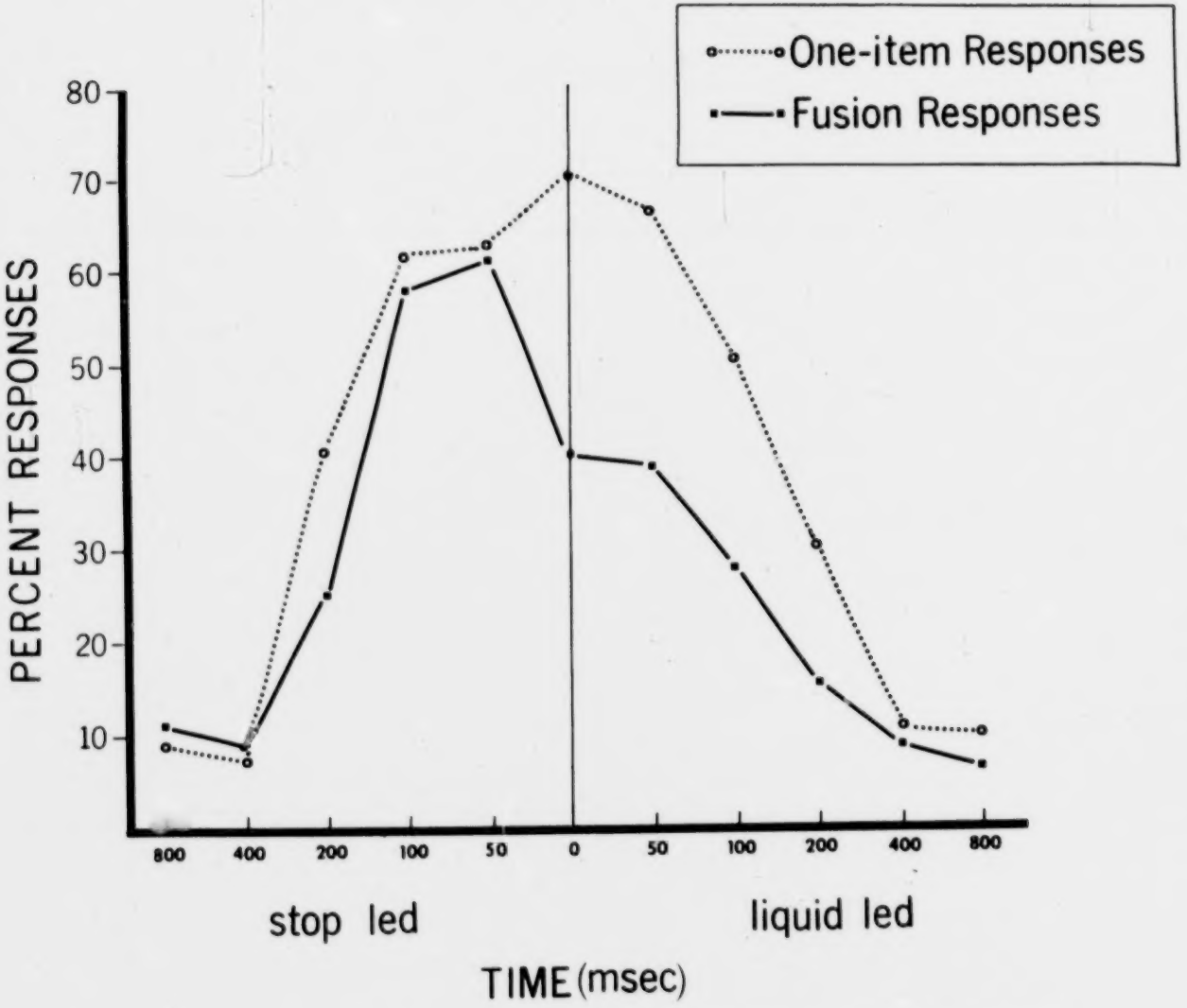


Figure 3: A comparison of fusion rate and the number of one-item responses.

the present study all stimuli were monosyllabic words, while previous studies (Day, 1968, 1970a, in preparation-b) used more complex disyllabic stimuli such as BANKET/LANKET. Longer stimuli may be less sensitive to lead time differences in the range between 100 and 200 msec. The difference between relatively simple and more complex stimuli may also have implications for the second principle. Day (1970a) has found that fusion rate for disyllabic pairs is nearly identical for liquid-leading pairs and stop-leading pairs, while the present study found differential fusion rates for these pairs.

Fusion rates for stop + /r/ and stop + /l/ stimuli were nearly identical. However, there was a disproportionately large number of stop + /l/ responses. For example, PAY/LAY yielded PLAY, while PAY/RAY also yielded PLAY on a large number of trials. In fact, /l/ was substituted for /r/ in the fusion response on nearly half of all trials in which stop + /r/ stimuli were fused. The reverse substitution was infrequent. Day (1968), Cutting and Day (1972), and Cutting (1973a, b) have reported this phenomenon and it is considered in more detail in later experiments when linguistic dimensions of the stimuli are varied. Fusion rates for the KICK set and the PAY set in the present study were comparable.

Overview. Timing is not a crucial factor in phonological fusion, since fusion continued to occur to a considerable extent at long lead times. Cutting (1972) has observed that no other auditory fusion occurs with lead times of greater than $\pm$ 25 msec. The observed insensitivity of phonological fusion to timing differences is congruent with the notion that it is a higher-level process.

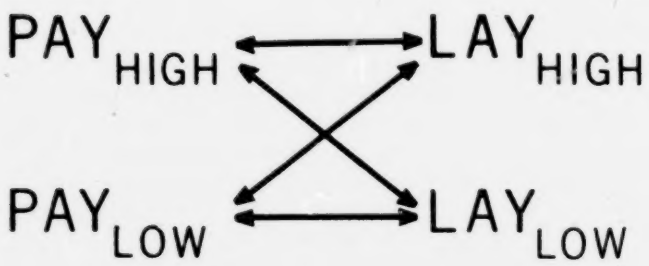
Experiments II and III: Pitch and Intensity

Insensitivity to parameters of pitch and intensity would lend further support to the classification of phonological fusion with the higher-level fusions.

Method. The same fusion sets as in Experiment I were used: the PAY set (PAY, RAY, LAY) and the KICK set (KICK, RICK, LICK). Two versions of each stimulus were synthesized for Experiment II, one at a relatively high fundamental frequency or pitch, and one at a relatively low pitch. All stimuli had a falling pitch contour. High-pitch stimuli began at 140 Hz and fell to a value of 120 Hz, while low-pitch stimuli began at 120 Hz and fell to a value of 100 Hz. The 20 Hz difference in pitch in this frequency range is equivalent to a difference of three notes on a musical scale. Thus, as shown at the top of Figure 4, the stimuli PAY and LAY were labeled: PAY-high (for "high" pitch), PAY-low, LAY-high, and LAY-low.

Two versions of each stimulus were also synthesized for Experiment III: one at a relatively high intensity (70 db SPL) and one at a relatively low intensity (55 db SPL). Both sets had the same low pitch used in Experiment II. The difference between the two intensities was 15 db, a difference which made the high-intensity stimuli perceptually 32 times more powerful than the low-intensity stimuli. The top of Figure 4 can again be used to display the stimuli; instead of "high" and "low" referring to pitch level, these terms refer to intensity level. Twelve subjects served in Experiment II and 12 different subjects in Experiment III.

STOP + LIQUID



"SAME" PAIRS

$PAY_{HIGH} + LAY_{HIGH}$

$PAY_{LOW} + LAY_{LOW}$

"DIFFERENT" PAIRS

$PAY_{HIGH} + LAY_{LOW}$

$PAY_{LOW} + LAY_{HIGH}$

Figure 4: Pairings of stop and liquid stimuli for the fusion task.

Dichotic pairs were assembled from possible combinations of fusible items within each set. As shown at the bottom of Figure 4, two pairs shared the same value of the target dimension: for example, PAY-high/LAY-high and PAY-low/LAY-low. The other two pairs had different values: PAY-high/LAY-low and PAY-low/LAY-high. Pairs that shared the same pitch or intensity ("same" pairs) and pairs that differed in pitch or intensity ("different" pairs) were presented in random order on the same tapes. Two tapes with different random orders were prepared for Experiment II, and two for Experiment III. Each tape contained 96 items: (2 sets of stimuli) x (2 stop/liquid pairs per set) x (4 pitch or intensity combinations) x (3 lead times) x (2 channel arrangements per pair). In order to measure differences in fusion rate as a function of the variation in pitch or intensity, a high overall fusion rate must be maintained for the "same" pairs. Experiment I showed that the 0 and ± 50 msec leads yielded substantial fusion rates, and hence these leads were used here. Within each experiment subjects listened to both stimulus tapes, with a brief rest between them.

Major results. Fusion occurred readily for all stimulus pairs. In Experiment II fusion rates were identical for pairs with the same pitch and for pairs with different pitches--50 percent each, as shown in Figure 5. In fact, fusion rates were within a few percentage points for all four combinations of pitch values.

The results of Experiment III were similar. Pairs with the same intensity and those with different intensities all fused at a rate of 36 percent, as shown in Figure 6. Again, there were no significant differences in fusion rate among the four combinations of intensity values.

Other results. a) Fusion rates were nearly identical for stop + /r/ and stop + /l/ stimuli. b) The /l/ substitutions were again frequent in both studies. When PAY/RAY fused, for example, PLAY responses occurred nearly 80 percent of the time in Experiment II and nearly 60 percent of the time in Experiment III. As reported earlier, /r/ substitutions were infrequent. c) Fusion rates for the PAY set and the KICK set were comparable within each study. d) Overall fusion rate for "same" pairs in Experiment III was lower than those in Experiment II, although this difference was not significant. A possible explanation for this difference in fusion rate may be that many of the Experiment III subjects were "low fusers." Elsewhere it has been shown that there is a bimodal distribution of subjects according to their fusion rates: some subjects fuse most of the time, while others rarely fuse (Day, 1970a). Individual differences in the present experiments are considered in the section on additional findings. e) The effect of relative onset time was the same in Experiments II and III as it was for the 0 and ± 50 msec leads in Experiment I. That is, fusions were more frequent at the + 50 msec lead than at 0 or - 50 msec. This same pattern of results occurred in all other studies in this series, and therefore will not be discussed again.

Overview. Within the range of values explored pitch and intensity were not important stimulus dimensions for phonological fusion. These are findings consistent with the notion that phonological fusion is a higher-level phonomenon.

Discussion: Nonlinguistic Experiments (I-III)

Phonological fusion is strikingly independent of various nonlinguistic characteristics of the to-be-fused stimuli. Time, pitch, and intensity differences were explored here. In recent studies, other nonlinguistic differences

Figure 5

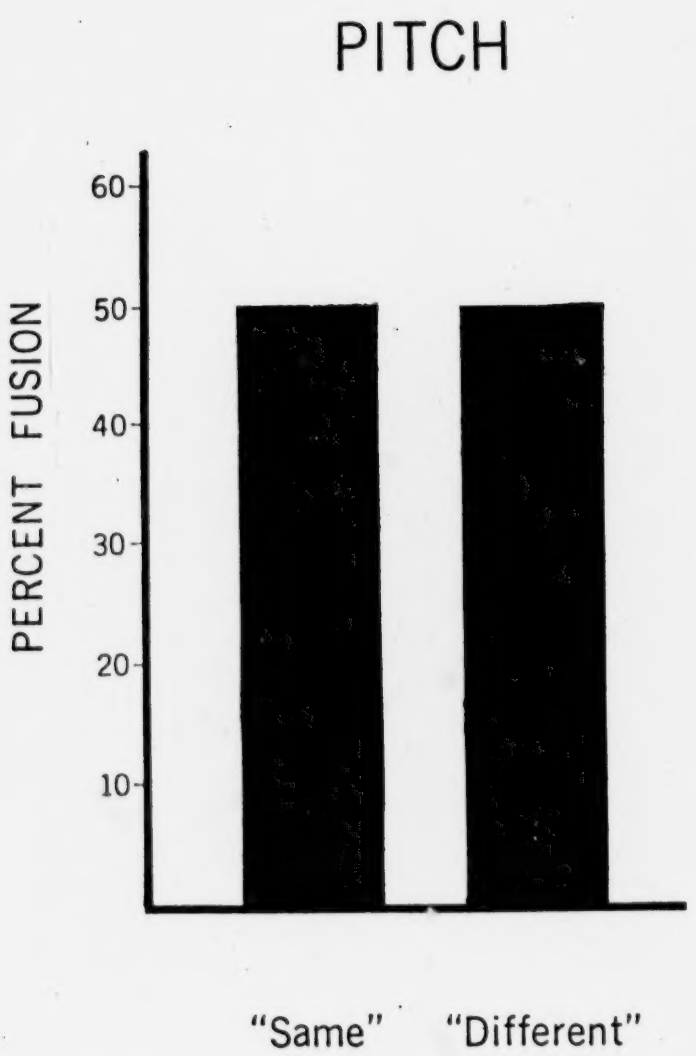


Figure 5: Results of the fusion task when pitch was varied.

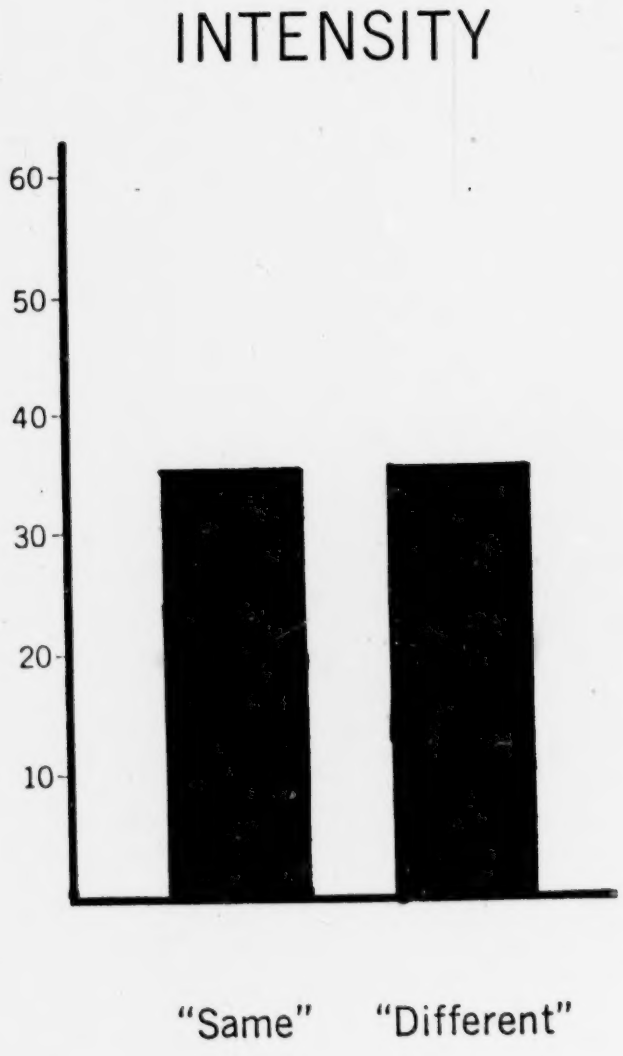


Figure 6: Results of the fusion task when intensity was varied.

were studied. Cutting (1973f) used the same paradigm shown in Figure 4 for another dimension, the apparent vocal tract size from which the fusible stimuli were spoken. Stimuli were synthesized to represent a relatively large vocal tract of a normal adult male (as used in all studies in this series) and the vocal tract of a midget or a small child. This manipulation again had no effect on fusion rate.

As a final demonstration of the independence of phonological fusion from nonlinguistic bonds, another study is relevant here. In previous studies only one parameter was varied at a time: for example, pitch, intensity, or vocal tract size. Cutting and Day (in preparation) varied all three dimensions: pitch, intensity, and vocal tract size. Results showed that even when all three dimensions of the stimuli were different for the two members of a pair fusion rate was unaffected. For example, PAY-low (pitch)-low (intensity)-large (vocal tract) fused as readily with LAY-high-high-small as it did with LAY-low-low-large.

Natural speech fusible items differ in nonlinguistic dimensions despite the care taken in uttering them. Cutting (1973a) compared the fusion rate for natural speech pairs with that of their more accurately controlled synthetic counterparts, and found that fusion rate was higher for synthetic items. The results of Experiment I-III have shown that relative onset time, pitch, and intensity do not have much influence on fusion rate. Perhaps the difference between the fusion rates of synthetic and natural speech pairs involves stimulus duration. The natural speech items for a given set often differed by as much as 50 msec in duration, while their synthetic counterparts were made to be equal. The effect of stimulus duration on phonological fusion has not yet been examined.

Six fusions reconsidered. Cutting (1972) described six different types of auditory fusion. The relative sensitivities of five fusions to time, pitch, and intensity variation in the to-be-fused stimuli were listed in Table 1. Table 2 adds phonological fusion to the higher-level fusions. Like phonetic feature

TABLE 2: An expanded version of Table 1 including phonological fusion among the other fusions.

	Timing	Pitch	Intensity
LOWER-LEVEL FUSIONS			
1. Sound Localization	<2 msec	<2 Hz	<2 db
2. Spectral Fusion	<5 msec	<2 Hz	*
3. Psychoacoustic Fusion	*	<2 Hz	*
HIGHER-LEVEL FUSIONS			
4. Phonetic Feature Fusion	25 msec	>20 Hz	>30 db
5. Chirp Fusion	25 msec	>20 Hz	>30 db
6. Phonological Fusion	150 msec	>20 Hz	>15 db

*Systematic data not available.

fusion and chirp fusion, phonological fusion is relatively insensitive to large stimulus differences in the three nonlinguistic dimensions. The lead time value of 150 msec was selected for phonological fusion, in part, as a mean value between 100 msec, where fusion occurred readily for the present stimuli, and 200 msec, where fusion occurred rarely, and in part because Day (in preparation-b) and Day and Cutting (1970) found that fusion of disyllabic stimuli occurred readily at 150 msec leads. The pitch value of 20 Hz matches the experimental values tested for other higher-level fusions, and preliminary work suggests that all higher-level fusions are tolerant of much greater pitch differences. The intensity value of 15 db reflects current knowledge since this is the largest interval studied to date.

The maximum relative onset value at which fusion occurs varies even within the group of higher-level fusions. Perhaps this differential sensitivity to timing can be explained in terms of the units which are fused. In both phonetic feature fusion and chirp fusion the units are phonetic: features of different phonemes are combined into a single new phoneme. In phonological fusion, however, the units are the phonemes themselves: phonemes from different stimuli are combined into a cluster. Since phonemes are made up of phonetic features, they are necessarily larger units than features. It follows that, since the units which are fused in phonological fusion are larger, the maximum relative onset value at which fusion occurs should also be larger. Indeed, phonetic feature fusion and chirp fusion begin to disintegrate with stimulus onset time differences of 25 msec, while phonological fusion rates remain high at differences of 100 or 150 msec.

Higher and lower levels reconsidered. From the data presented in Table 2, a process model is proposed describing the differences between higher- and lower-level fusions, as shown in Figure 7. For the sake of simplicity, the perceptual system was divided into two parts: a higher-level processor and a lower-level processor. Two dichotic stimuli, Stimulus A and Stimulus B, necessarily enter the system by way of the lower-level processor and are sent upwards in the system. Two experimental situations have been considered: one in which there is no variation in nonlinguistic stimulus dimensions (cases 1 and 2), and one in which such variation does occur (cases 3 and 4). In the no-variation conditions, higher- and lower-level fusions cannot be distinguished, since fusion occurs readily for both. This situation is represented for cases 1 and 2 by the solid lines and arrows. In the variation condition differences occur: higher-level fusions are not disrupted by nonlinguistic stimulus variation, while lower-level fusions are disrupted. The information in the stimuli must therefore be extracted at different levels for the two types of fusion. In higher-level fusions, the stimuli are combined into a single percept in the higher-level processor (case 3), and since nonlinguistic stimulus variation has little effect on fusion rate higher-level fusions may take place exclusively in the higher-level processor (hence the dashed lines for case 1). Lower-level fusions, on the other hand, occur only in the lower-level processor since fusion is disrupted when nonlinguistic variation occurs (hence the broken arrows in case 4).

Higher-level fusions are influenced by linguistic variables and lower-level fusions are influenced by nonlinguistic variables. One corollary to this statement is that higher-level fusions occur only for speech stimuli, whereas lower-level fusions occur for both speech and nonspeech. A second corollary is that the higher-level processor shown in Figure 7 is basically a speech, or language, processor, whereas the lower-level processor is concerned with auditory aspects

Figure 7

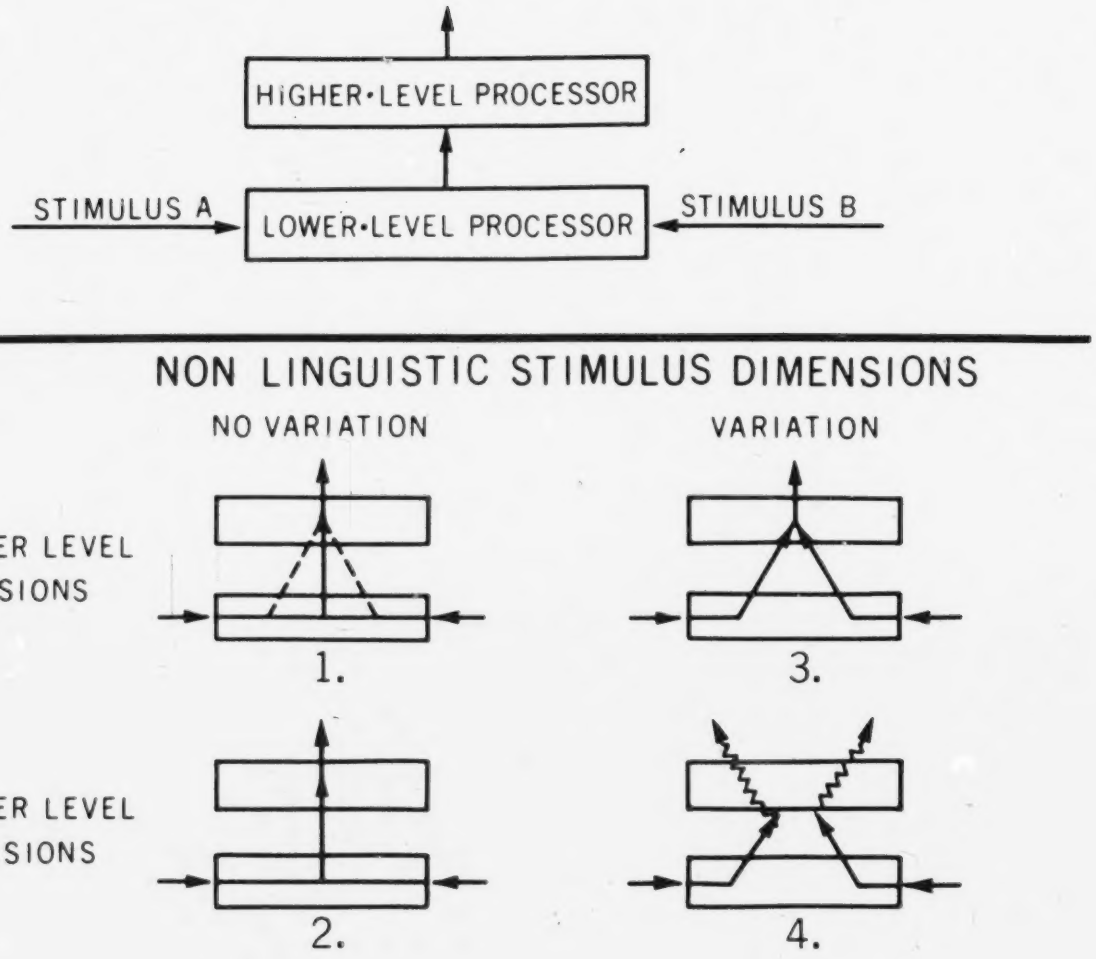


Figure 7: A process model comparing higher- and lower-level fusions.

of the signal. The two processors, then, are necessarily hierarchical: a signal which has linguistic dimensions must also have nonlinguistic dimensions, such as pitch and intensity. A signal which has nonlinguistic dimensions, on the other hand, need not necessarily have linguistic dimensions.

Since phonological fusion is governed by higher-level rules, the rules of language, the remaining experiments were designed to study the specific linguistic levels at which fusion takes place.

II. LINGUISTIC DIMENSIONS

Three linguistic levels were considered in the remaining experiments: the level of semantics, the level of the phoneme, and the level of acoustic structure as it pertains to language. The phoneme level might appear to be the primary linguistic level at which phonological fusion occurs, since it is phonemes which are fused into clusters. However, this need not be the case. Therefore the two other levels were chosen to be higher (semantics) and lower (acoustics) than the phoneme level. The experiments begin at the semantic level and move "downward" to the phoneme and acoustic levels.

Semantic Level

Day (1968) found that semantic cues at the word level influenced fusion rate. Fusion rates were higher when the fused outcomes were real words than when they were nonwords (PAHDUCT/RAHDUCT⟶PRODUCT vs. PAHLOW/RAHLOW⟶PRAHLOW). Nonword fusions did occur (GORIGIN/LORIGIN⟶GLORIGIN) although at a reduced rate.

Experiment IV: Sentence Context

The present experiment was designed to observe the effects of semantic cues at the sentence level on fusion rate. Since Experiments I-III found that /l/ was frequently substituted for /r/ in fusion responses, the present study was also designed to observe the effect of semantic context on /l/ substitutions.

Method. Two sets of stimuli were used, the PAY set (PAY, RAY, LAY) and the GO set (GO, ROW, LOW). Fusible pairs were presented in isolation, as in previous experiments, and imbedded in sentence contexts. The PAY set appeared in the contexts THE TRUMPETER ____S FOR US and THE MINISTER ____S FOR US, while the GO set appeared in THE COALS ARE ____ING AGAIN and THE TREES ARE ____ING AGAIN. These sentences were made into dichotic pairs such that THE TRUMPETER PAYS FOR US, for example, was presented to one ear and THE TRUMPETER LAYS FOR US to the other. Fusible pairs appeared in both semantically appropriate and inappropriate contexts. Appropriateness was determined in a rating experiment with separate subjects where pairs of sentences such as THE TRUMPETER PLAYS FOR US and THE TRUMPETER PRAYS FOR US were presented in written form and subjects were asked to judge which was most "meaningful." The "meaningful" ratings were taken as a measure of semantic appropriateness, and the most appropriate member of each pair is shown in a box in Figure 8. Additional details are given in Appendix C.

Two tapes were prepared, one with fusible targets imbedded in sentences and the other with the target pairs presented in isolation. The sentence tape consisted of 64 items: (4 sentence frames) x (2 stop/liquid pairs per set) x (2

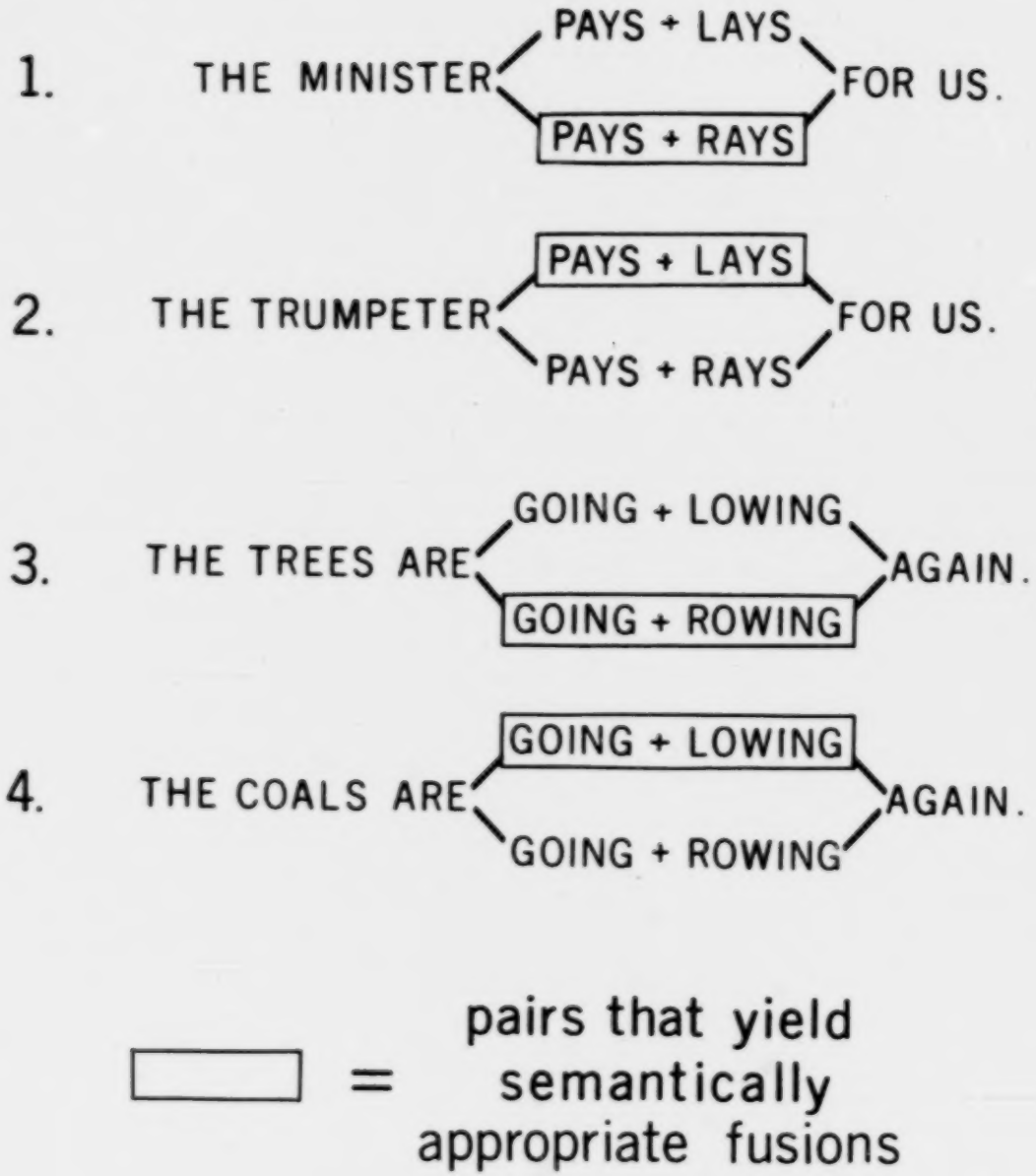


Figure 8: Fusible pairs imbedded in sentence frames.

channel arrangements) x (4 observations per sentence). Dichotic sentence pairs were presented at a simultaneous onset with 12 seconds between trials. Subjects wrote down the entire sentence. The no-sentence tape also had 64 pairs: (2 sets of stimuli) x (2 stop/liquid pairs per set) x (2 channel arrangements) x (8 observations per pair). Again, only the simultaneous onset time was used, but the intertrial interval was four seconds. Subjects wrote down 'what they heard,' one word or two words, acceptable words or nonsense. Half of the 16 subjects listened first to the sentence tape and then to the no-sentence tape, while the others listened in reverse order.

Major results. Fusion rate was significantly higher for the sentence condition than for the no-sentence condition, as shown in Figure 9. All subjects showed this trend ($z = 3.8$, $p<.001$). Fusion rate was 85 percent for sentence trials and it was 65 percent for no-sentence trials, a rate comparable to that found in previous studies.

Other results. a) The order in which subjects listened to the sentence and no-sentence tapes was not a significant factor. b) Fusion rates were comparable for stop + /r/ and stop + /l/ items, as well as, c) for both sets of stimuli in each condition.

Figure 10 shows the percent responses in each sentence frame. Stop + /l/ responses dominated all sentence contexts even when they were semantically inappropriate. For sentence frame 1, THE MINISTER PLAYS FOR US occurred on 74 percent of all trials. Certainly, the minister PLAYING is semantically less likely than PRAYING even in today's society of changing roles. Likewise, in sentence frame 3, THE TREES ARE GLOWING AGAIN occurred on 83 percent of all trials, despite the fact that it was not judged to be very appropriate. Sentence frames 2 and 4 yielded a more predictable set of results: in both cases stop + /l/ fusions were judged semantically appropriate and these fusions vere very frequent. "Other" responses for all four sentence frames were primarily responses in which only the stop stimulus was reported; for example, THE MINISTER PAYS FOR US. Pairs of stimuli in the no-sentence condition yielded similar liquid substitution results: for example, when PAY/RAY fused, PLAY responses were given 85 percent of the time. The reverse substitution rarely occurred.

The present experiment showed that meaning at the sentence level could not account for the /l/-substitution effect. Relative frequency of occurrence of the fused words cannot account for them either: GLOWING, for example, is much less frequent than GROWING (Thorndike and Lorge, 1944; Carroll, Davies, and Richman, 1971). The relative frequency of these clusters in English fared little better as a predictor of the particular fusion response. In fact, stop + /r/ clusters occur almost twice as frequently as stop + /l/ (Day, 1968). Meaning at the word level may provide a clue to /l/ substitutions. Day (1968) found that, while subjects usually reported hearing GROCERY when given GOCERY/ ROCERY, sometimes they reported hearing GLOCERY, a nonword. In the present series of experiments, both the stop + /r/ and stop + /l/ fusions for a given set were acceptable English words. Day (1968, 1970a), on the other hand, chose stimuli that could fuse meaningfully in only one direction. She found that PAHDUCT/RAHDUCT⟶PRODUCT and not PLODUCT, and that GEEDY/REEDY⟶GREEDY not GLEEDY. Such results suggest that meaning at the word level can override the /l/-substitution effect. This effect is considered again in Experiment VI and in the section on additional findings.

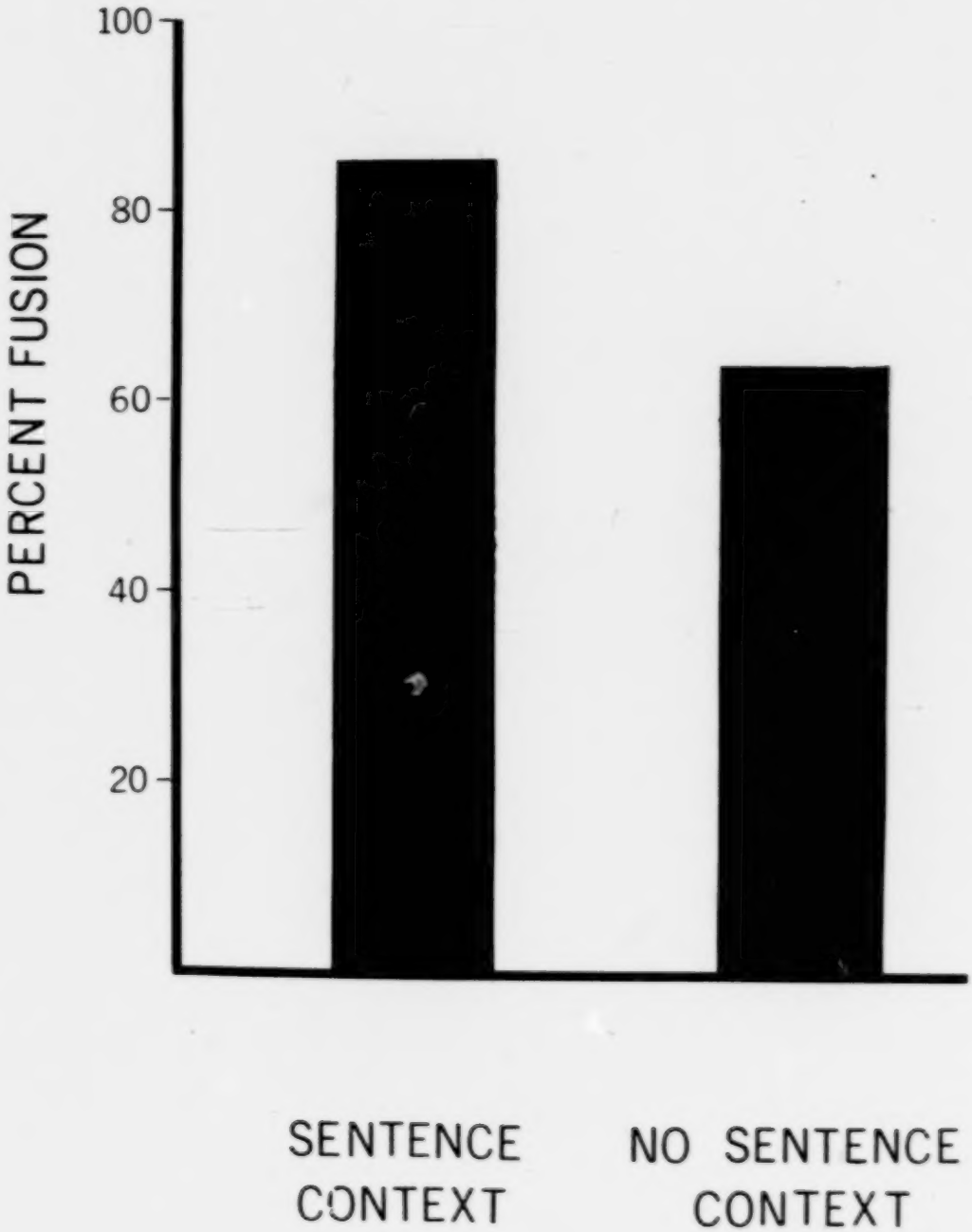


Figure 9: Results of the fusion task for sentence and no-sentence context conditions.

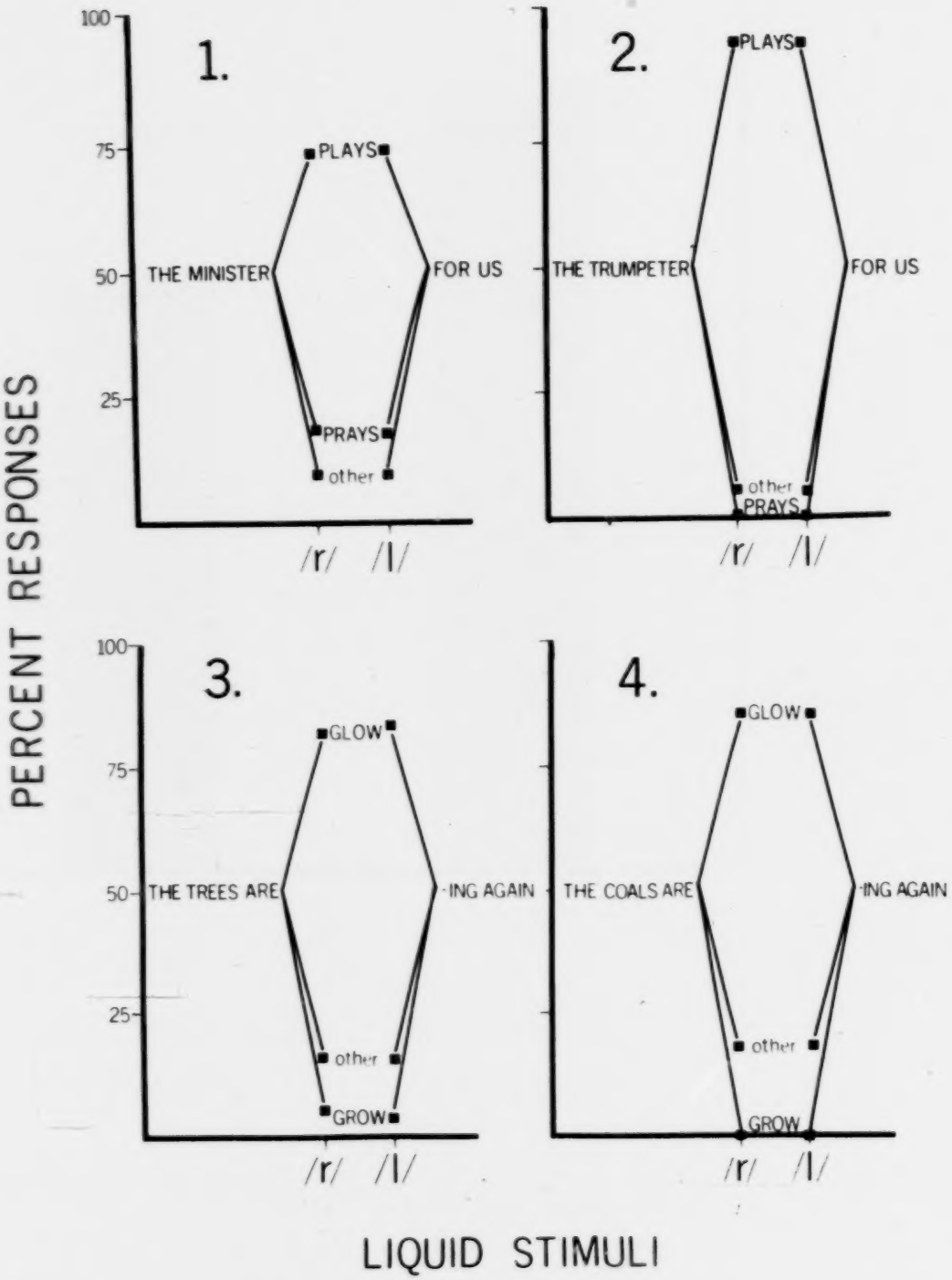


Figure 10: Percent responses for all four sentence frames.

Overview. Fusion rate was increased by higher-level, semantic cues which were outside the fusible stimuli themselves. However, sentence context cannot fully account for fusion since fusion rates continued to be high in the no-sentence condition.

Semantic context did not affect the type of fusion responses which subjects reported. Stop + /l/ fusions occurred even when the stop + /r/ responses would have been semantically more appropriate. However, the results of other experiments suggest that semantic cues at the word level can override the /l/-substitution effect.

Phoneme Level

Phonological fusion occurs when phonemes from different dichotic stimuli are perceived as a cluster. Experiments V and VI examined the phonemic components, the stop and the liquid, to assess their importance in the fusion phenomenon.

Experiment V: Stops and Fricatives

In Experiments I-IV stop consonants served as the first phoneme of the to-be-fused consonant-consonant-vowel cluster. The present study compared the fusibility of stop/liquid and fricative/liquid pairs. Both initial clusters occur very frequently in English, but that fact does not necessarily imply that they fuse equally well.

Method. In addition to the BED set (BED, RED, LED) and the GO set (GO, ROW, LOW) the fricative stimuli FED and FOE were synthesized.[4] The fricative stimuli were identical to the stop stimuli in duration, pitch, and intensity, and differed only in the acoustic structure of the first 100 msec as shown in Figure 11. Appropriate frication for the phoneme /f/ was substituted for the formant transitions and initial vowel segments of BED and GO. A given liquid stimulus such as LED was paired with both a stop (BED/LED) and a fricative (FED/LED). All stimuli and possible fusions were English words or names: BED/RED⟶BREAD, BED/LED⟶BLED, FED/RED⟶FRED, FED/LED⟶FLED, GO/ROW⟶GROW, GO/LOW⟶GLOW, FOE/ROW⟶FRO, FOE/LOW⟶FLOW.

One tape was prepared with stop/liquid pairs and another tape with fricative/liquid pairs. Each tape consisted of 120 dichotic trials: (2 sets of stimuli) x (2 consonant/liquid pairs per set) x (3 lead times) x (2 channel arrangements) x (5 observations per pair). Twelve subjects listened to both tapes: half listened first to the fricative/liquid stimuli and then to the stop/liquid stimuli, while the others listened in the reverse order.

Major results. Fusion occurred much more readily for the stop/liquid pair than the fricative/liquid pairs. Fusion rates were 57 percent and 18 percent respectively, as shown in Figure 12. This 3:1 ratio was highly significant, with all subjects showing greater fusion rates for stop/liquid items (z = 3.18,

---

[4] The fricative /f/ was chosen because it is the only fricative in English that clusters with both /r/ and /l/ in initial position.

Figure 11

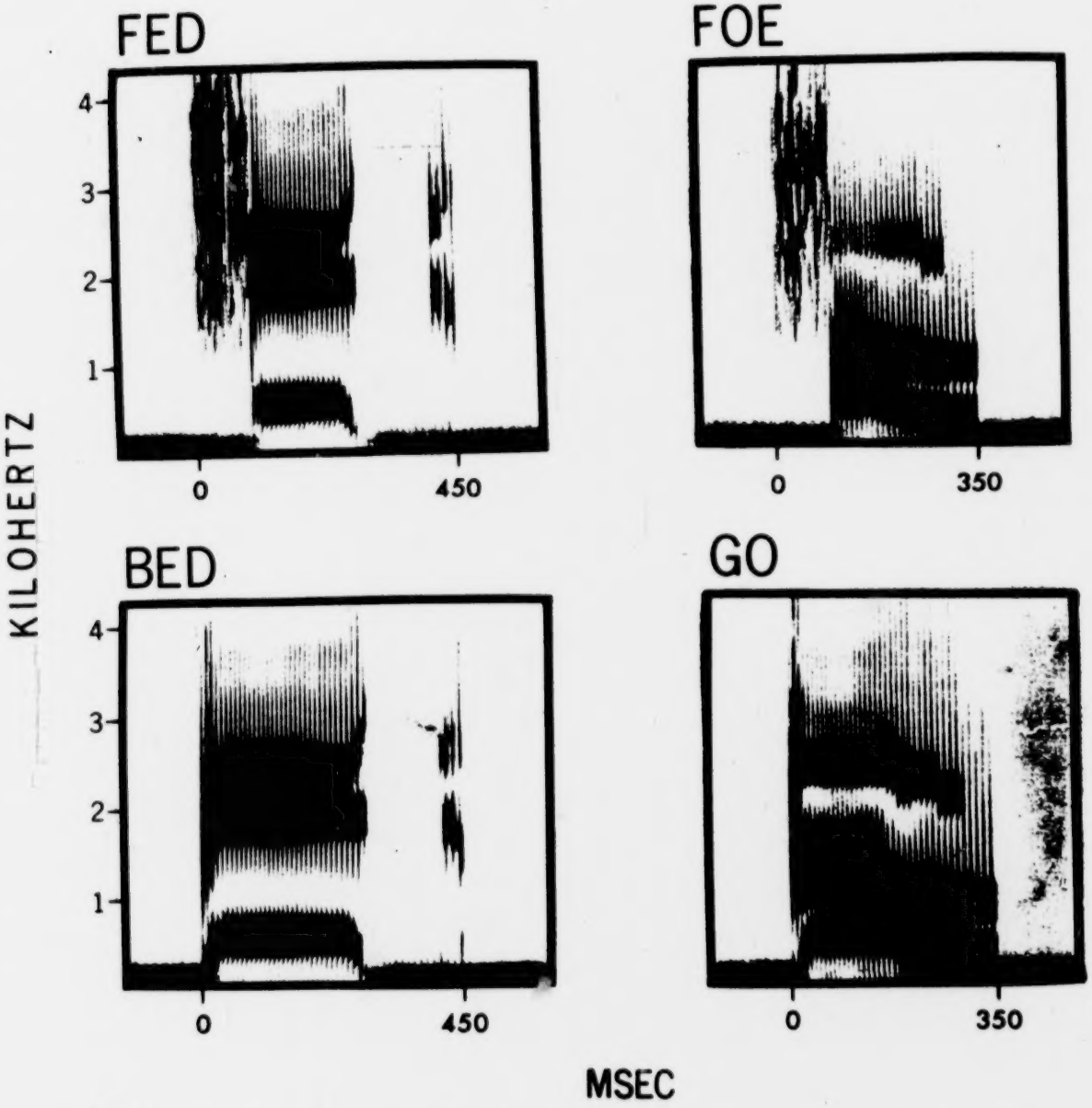


Figure 11: Spectrograms of fricative-initial and stop-initial stimuli.

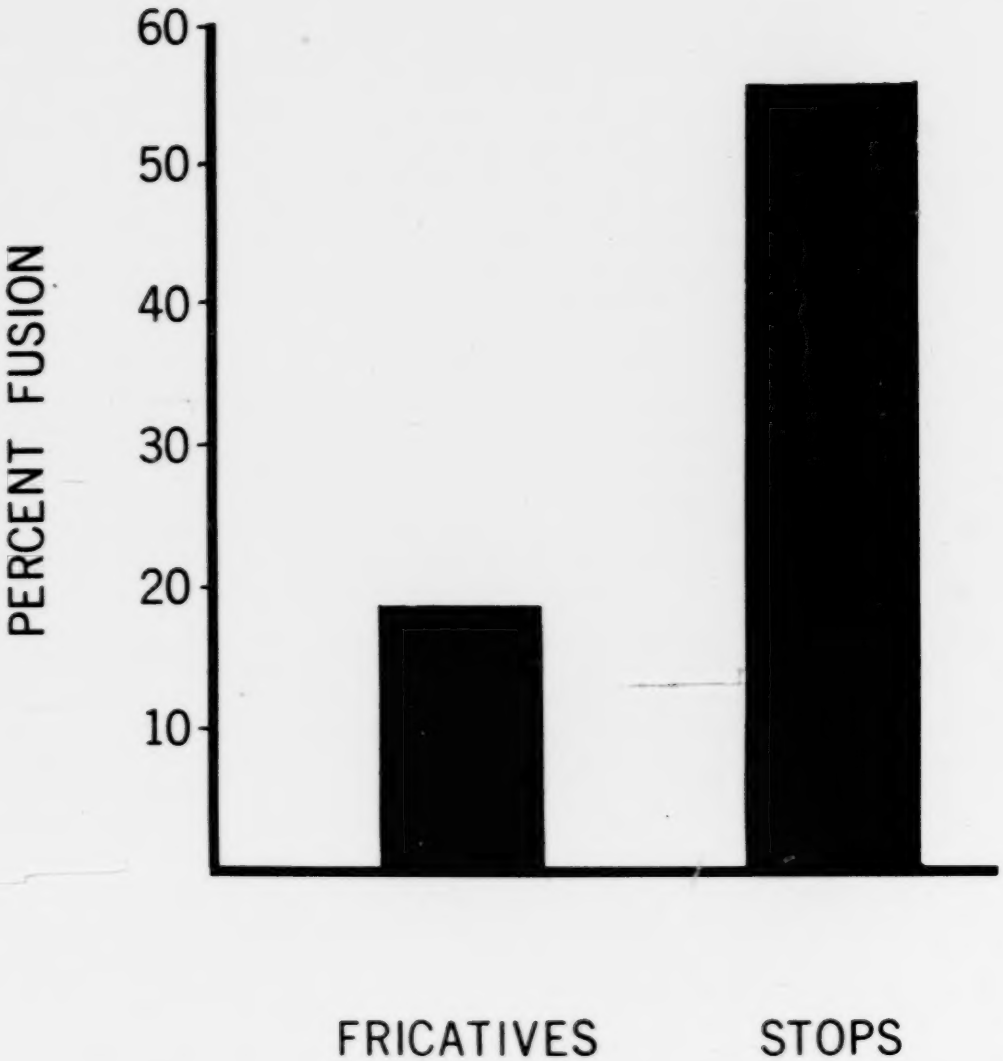


Figure 12: Results of the fusion task for fricative/liquid and stop/liquid pairs.

$p<.001$). Fusion rate differences cannot be accounted for by the relative frequency of the possible fusion responses in English: for example, BLED and GLOW are much less common than FLED and FLOW (Carroll, Davies, and Richman, 1971). Furthermore, initial /f/ + liquid clusters occur at about the same frequency as initial /b/ + liquid clusters in English, and considerably more frequently than initial /g/ + liquid clusters (Hultzén, Allen, and Miron, 1964; Denes, 1965).

Perhaps differences in fusion rates for the two classes of consonants can be accounted for on other grounds, such as relative encodedness. Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) have defined encodedness as the general amount of acoustic restructuring that a phoneme undergoes in different contexts. Stop consonants are highly encoded, whereas fricatives are only moderately encoded. Perhaps highly encoded phonemes combine more easily with other phonemes to produce a cluster.

Other results. a) The order in which subjects listened to the stop and fricative tapes was not a significant factor. b) Fusion rates for stop + /r/ and stop + /l/ stimuli were again within a few percentage points. Fricative + /r/ and fricative + /l/ fusion rates were also comparable. c) The /l/ substitutions were frequent for stop/liquid pairs, but not for fricative/liquid pairs. In fact, /r/ substitutions occurred on 64 percent of all trials in which fricative + /l/ stimuli fused, while corresponding /l/ substitutions were rare. No explanation is apparent for this reversal in liquid substitutions for fricative/liquid stimuli.

A second type of substitution also occurred. Fricative/liquid stimuli did not always yield fricative + liquid responses; for example, FED/RED⟶BRED. In fact, about 70 percent of all fricative/liquid pair fusions were actually stop + liquid responses. Stop-for-fricative substitutions were not the result of poor fricative stimuli, since subjects identified them in isolation on the diotic test with a high degree of accuracy (see Appendix B). Instead, these substitutions appear to be an extension of the differences in fusibility between the stops and fricatives.

Overview. When the first consonant in the to-be-fused cluster was a stop, fusion rate was high, but when it was a fricative, fusion rate dropped.

Experiment VI: Liquids and Semivowels

The present study examined the second consonant in the to-be-fused cluster. Stimuli beginning with semivowels (that is, /w/ and /y/) were prepared to see whether they would fuse as readily as the liquids, and to see whether the /l/-substitution effect would be extended to /w/ and /y/.

Method. Two sets of stimuli were used: the KICK set (KICK, RICK, LICK, WICK) and the COO set (COO, RUE, LIEU, YOU).[5] Liquid and semivowel stimuli

[5] The only stop consonant that clusters with all liquids and semivowels in English is /k/; yet /ky/ occurs only before the vowel /u/, while /kw/ does not occur before /u/. Thus, it was necessary to synthesize two sets of stimuli: one for /r, l, w/ and the other for /r, l, y/. Note that LIEU could also be represented as LOU.

within the same set were identical in all respects except for the direction and slope of the second and third formant (F2 and F3) transitions, as shown in Figure 13. These are the cues that distinguish all liquids and semivowels (O'Connor et al., 1957; Lisker, 1957).

Figure 14 shows the stimuli and the possible fusion responses for both sets. All were words or names common in English. A tape was prepared with 108 dichotic items: (2 sets of stimuli) x (3 liquid and semivowel stimuli per set) x (3 lead times) x (2 channel arrangements) x (3 observations per pair). Twelve subjects listened to two passes through the tape, reversing headphones after the first pass.

Major results. Fusion rate was comparable for pairs within a particular stimulus set as shown in Figure 15. KICK/RICK, KICK/LICK, and KICK/WICK pairs all fused at an average of 70 percent; while COO/RUE, COO/LIEU, and COO/YOU all fused at an average of 42 percent. There were no significant differences within each set.

Other results. Regardless of which stimuli were presented, most responses were stop + /l/; /l/ was substituted for /r/, as in previous studies, and it was also substituted for /w/ and /y/. CLICK and CLUE responses occurred in 89 percent of all trials in which fusions occurred. Again, word frequency of the possible fusions cannot account for the substitutions. For example, QUICK is much more common than CLICK, and CREW is more common than CLUE (Carroll, Davies, and Richman, 1971). Nevertheless, the /l/ substitutions for both sets of stimuli yielded relatively common English words. The data of Day (1968) suggest that when /l/ substitutions do not yield acceptable words, they occur considerably less often.

The KICK set fused more than the COO set. Word frequency cannot account for this difference. Other possible causes of differential fusion rates among stimulus sets include cluster frequency, phonetic differences in the stop stimuli, and phonetic differences in the vowels (see Cutting, 1973a).

Overview. The role of the second consonant in the to-be-fused cluster is less clear than that of the first consonant. For the present stimuli, fusion occurred equally well for all stop/liquid and stop/semivowel pairs, yet all pairs tended to yield a stop + /l/ response.

Acoustic Level

Linguistic cues at the sentence and phoneme levels have been shown to affect fusion rate (Experiments IV-VI). Perhaps linguistic cues at the level of acoustic structure are also important. Since the liquid is perceptually interpolated between the stop and the vowel, one key to fusion may lie in its acoustic structure. Experiments VII-IX examined various aspects of the acoustic structure of liquids.

Experiment VII: Liquid Transitions

Experiment VI showed that for the present sets of stimuli, liquids (/r, l/) and semivowels (/w, y/) tended to yield stop + /l/ fusions. Since /l/, /w/, and /y/ have falling F3 transitions while /r/ has a rising F3, and since /r/, /l/,

Figure 13

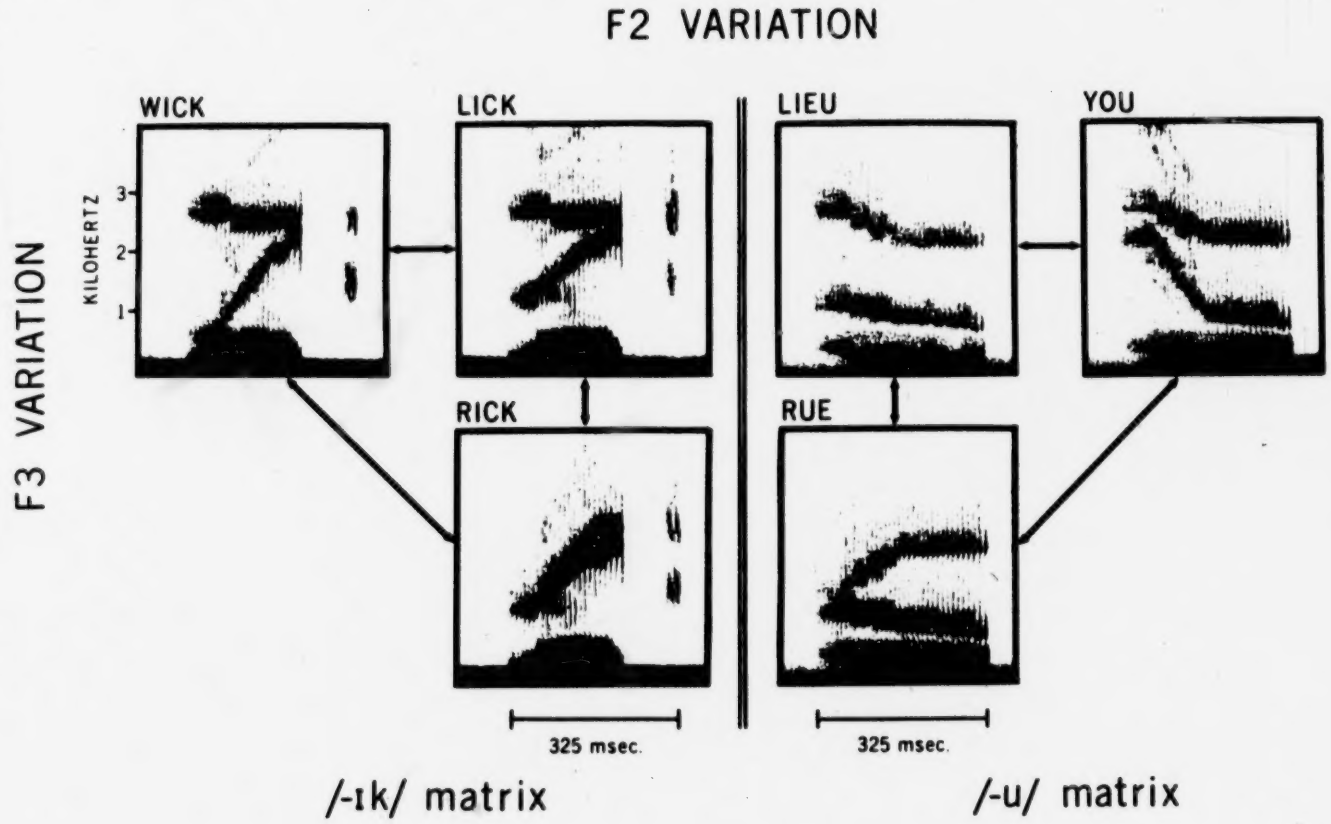


Figure 13: Spectrograms of liquid-initial and semivowel-initial stimuli.

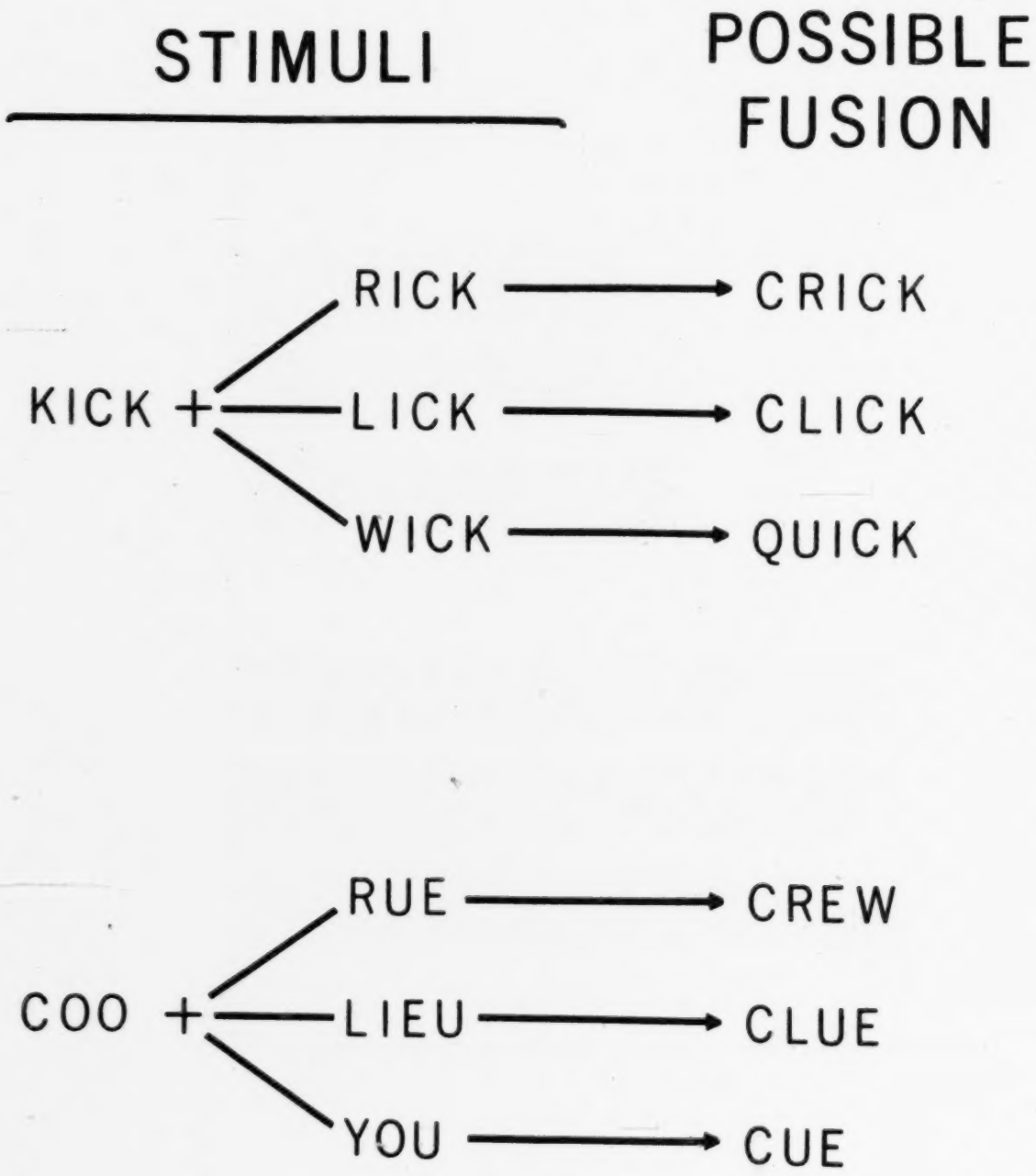


Figure 14: Stimuli and possible fusions when the liquid-initial stimulus is varied.

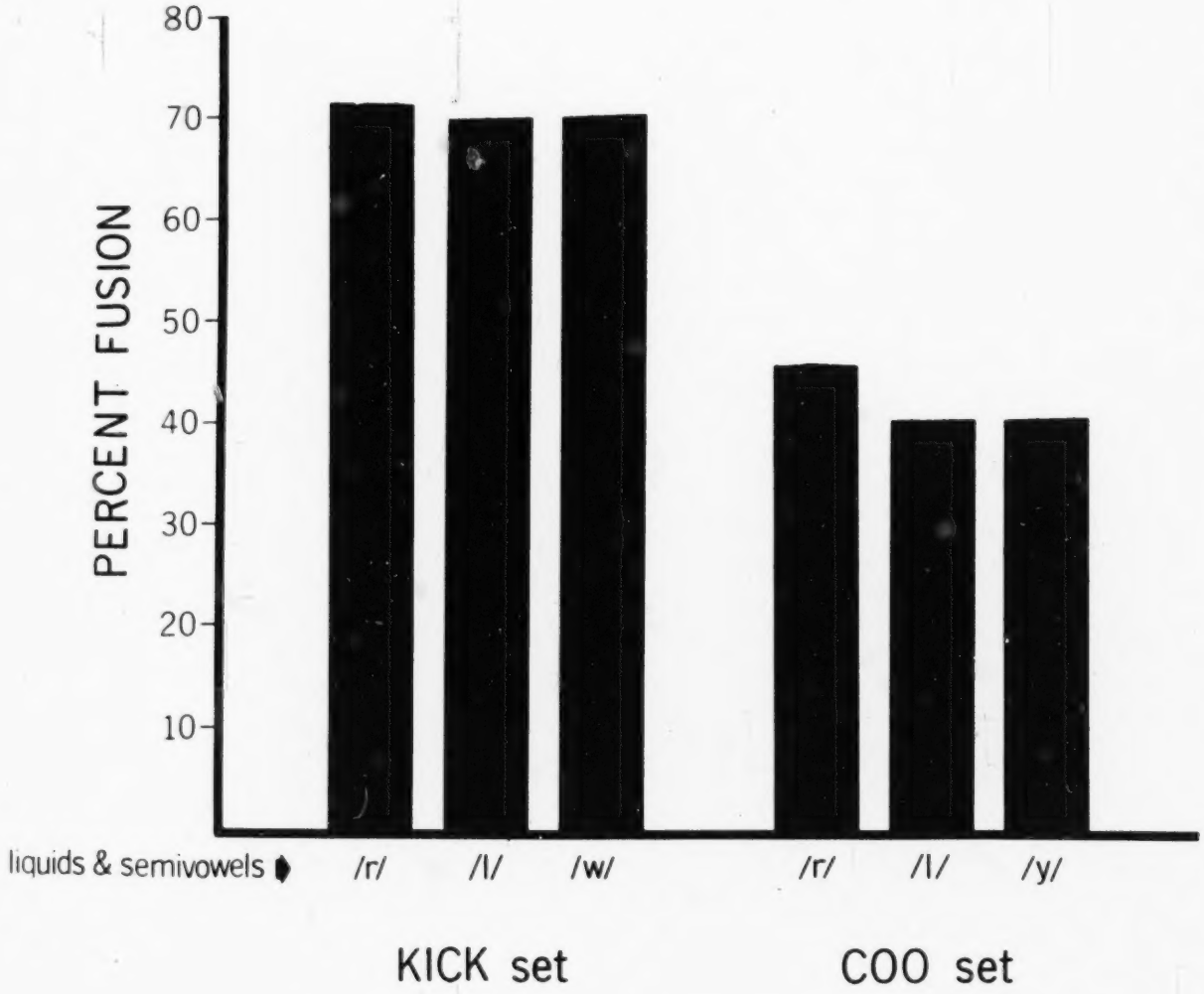


Figure 15: Results of the fusion task for stop/liquid and stop/semivowel pairs.

and /w/ have rising F2 transitions while /y/ has a falling F2 (see Figure 13), it appears that any combination of rising and falling F2 and F3 transitions is sufficient for stop + /l/ fusions to occur. If any combination is sufficient for fusion, perhaps no transitions are needed at all. For example, PAY/AY, a pair without any liquid transitions, might also yield stop + /l/ fusions. The present study varied the slope of the liquid transitions to determine how much, if any, transition was necessary for fusion to occur, and to confirm that fusion is not a response bias.

Method. The PAY set and the KICK set were expanded to include five stimuli, one stop stimulus and four stimuli which formed a liquid-to-vowel continuum, as shown in Figure 16. At one end of the continuum, Stimulus 1 had full liquid transitions in all formants as found in LAY and LICK. At the other end of the continuum, Stimulus 4 had the same duration but began with a steady-state vowel, AY and ICK. Between the extremes were Stimuli 2 and 3 which had intermediate amounts of formant transitions. Equal increments of acoustic change occurred between successive stimuli.

Sixteen subjects served in two tasks, identification of the liquid-to-vowel stimuli in isolation and dichotic fusion. Since the results of the identification task were highly relevant to the fusion task, those data are discussed first.

Task 1: Identification of the liquid-to-vowel stimuli

Tapes and procedure. A diotic identification tape of 80 randomized liquid-to-vowel items was prepared: (2 sets of stimuli) x (4 stimuli per array) x (10 observations per stimulus). There was a three-second interval between each item. Subjects wrote down the single item that they heard presented on each trial.

Results. Stimuli 1 and 2 were identified as beginning with /l/ on 88 percent of all trials, as shown in the top part of Figure 17. Stimuli 3 and 4 were identified as beginning with /l/ on only 8 percent of all trials. All subjects showed this quantal trend ($z = 3.8$, $p<.001$). These results demonstrate the well-known fact of categorical perception in certain speech sounds (see Liberman, 1957; Pisoni, 1971). Equal amounts of change along a physical dimension were not perceived as equally spaced, but instead were perceived in groups with a distinct boundary between Stimulus 2 and Stimulus 3. There was no difference between the LAY-to-AY and LICK-to-ICK stimulus arrays.

Task 2: Fusion

Tapes and procedure. Dichotic items were constructed by pairing the stop stimuli with all items in the liquid-to-vowel arrays. The tape consisted of 96 pairs: (2 sets of stimuli) x (4 stimuli per array) x (3 lead times) x (2 channel arrangements) x (2 observations per pair). Subjects listened to two passes through the tape, reversing headphones after the first pass. As usual they wrote their responses, indicating what they heard.

Results. Fusion occurred at a rate of 52 percent for pairs containing Stimulus 1 or Stimulus 2, the stimuli which had been identified as beginning with a liquid. Other pairs yielded only 6 percent fusions, as shown in Figure 17. No subject deviated markedly from the group data.

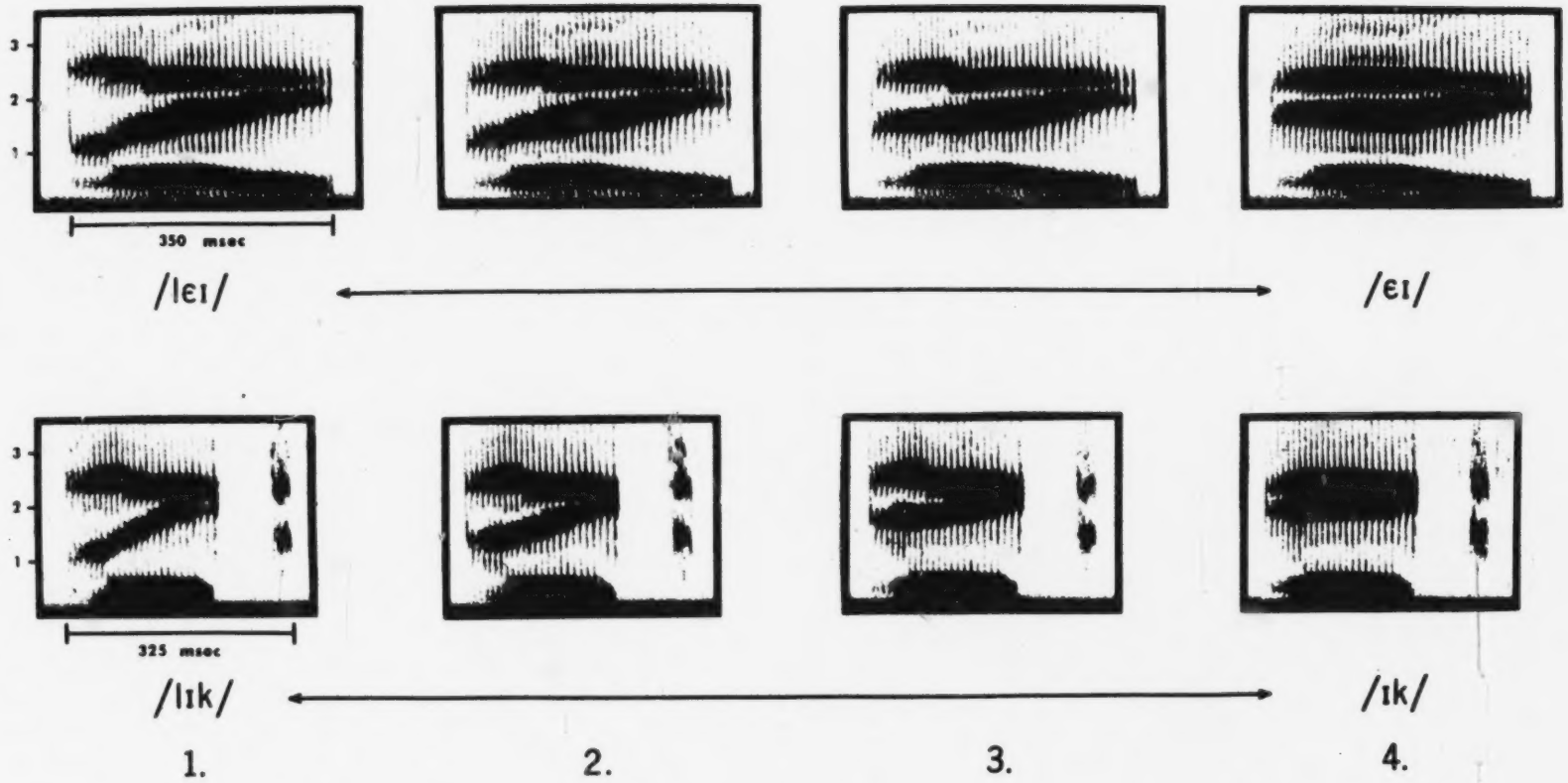


Figure 16

Figure 16: Spectrograms of liquid-initial to vowel-initial stimulus arrays.

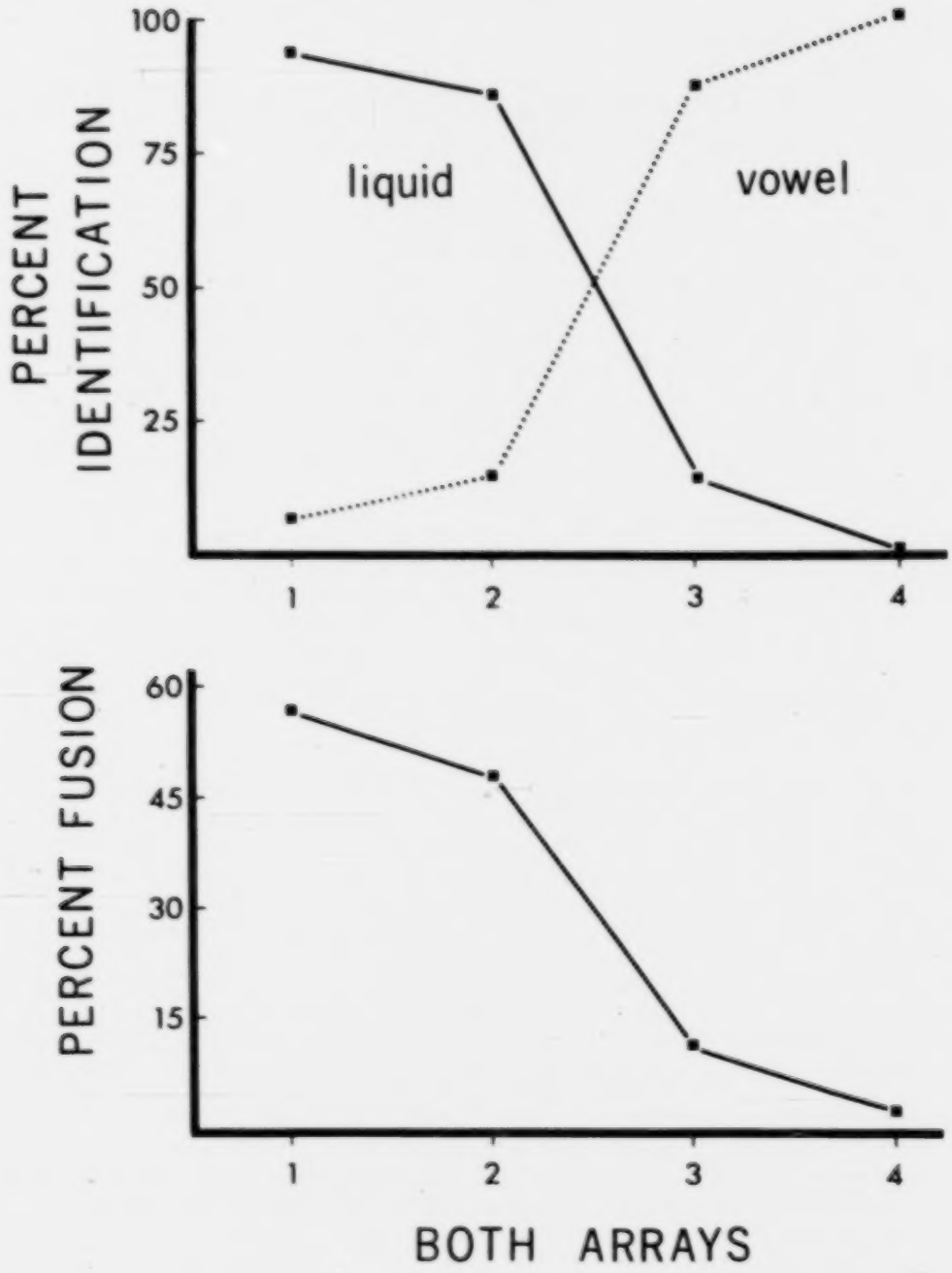


Figure 17: Results of identification and fusion tasks involving liquid-to-vowel stimulus arrays.

Overview: Liquid-like transitions were necessary for the phonological fusion of dichotic stop/liquid stimuli. Fusion occurred in direct proportion to the extent that "liquid-like" items were indeed perceived as liquids in isolation.

Experiment VIII: Degraded Liquids

Experiment VII showed that transitions in the liquid stimulus were necessary for fusion to occur. The present experiment was designed to determine which formant transition (or combination of transitions) is necessary for fusion.

Method. The PAY set and the KICK set were again used. Liquid stimuli appeared in many forms. Some were degraded in that certain formants were omitted from their acoustic structure. Figure 18 shows the component parts of the liquid stimuli. There were two possible third formants, that for /r/ and that for /l/. F3/r/ represents the third formant of the /r/ stimuli, while F3/l/ represents the third formant of the /l/ stimuli. All possible combinations of F1, F2, F3/r/, and F3/l/ were used. Eleven liquid-like stimuli resulted in each set: 2 three-formant stimuli identical to those used in previous studies, 5 two-formant stimuli, and 4 one-formant stimuli. Two-formant stimuli were F1+F2, F1+F3/r/, F1+F3/l/, F2+F3/r/, and F2+F3/l/. One-formant stimuli were F1, F2, F3/r/, and F3/l/.

Each of the 11 liquid-like stimuli was paired with its appropriate stop stimulus. In addition two control pairs were constructed per set: one pair was a stop/stop pair (PAY/PAY and KICK/KICK), and the other was a stop presented to one ear and nothing to the other (PAY/--- and KICK/---). No fusion responses should occur for control pairs if fusion occurs only for pairs containing liquid-like stimuli. A dichotic tape of 156 items was constructed: (2 sets of stimuli) x (13 pairs per set) x (3 lead times) x (2 channel arrangements per pair). Twelve subjects listened to one pass through the tape.

Major results. There were two general fusion rates: most experimental pairs fused at a rate of about 55 percent, while a few pairs fused at a considerably lower rate, as shown in Figure 19. Pairs which rarely fused contained liquid-like stimuli with only F1 or F3/l/ and no other formant. Figure 18 shows that these stimuli lacked formant transitions in the mid-frequency range (1000-2000 Hz), while all other liquid-like stimuli had transitions in this region (either F2 or F3/r/).

Stop + three-formant liquid stimuli fused at rates comparable to previous studies--54 percent, with no significant difference between stop + /r/ and stop + /l/. Stop + two-formant liquid pairs fused at a rate of 52 percent, with the exception of the stop + F1,3/l/ case, where the fusion level was only 23 percent. All subjects showed this drop in fusion rate ($z = 3.18$, $p<.001$). Stop + one-formant liquid pairs also showed high and low fusion rates. Stop + F2 and stop + F3/r/ liquid pairs fused at a rate of 62 percent while stop + F1 and stop + F3/l/ liquid pairs fused at a rate of only 18 percent. Again, all subjects showed these quantal differences in fusion rate ($z = 3.18$, $p<.001$).

Other results. a) As in previous studies, stop + /l/ responses occurred on more fused trials than stop + /r/. b) Stop/stop pairs yielded few stop + /l/ responses. Such responses would be "false fusions" since the subject would be reporting a liquid which has not, in fact, been presented (Day, 1968).

LIQUID STIMULI

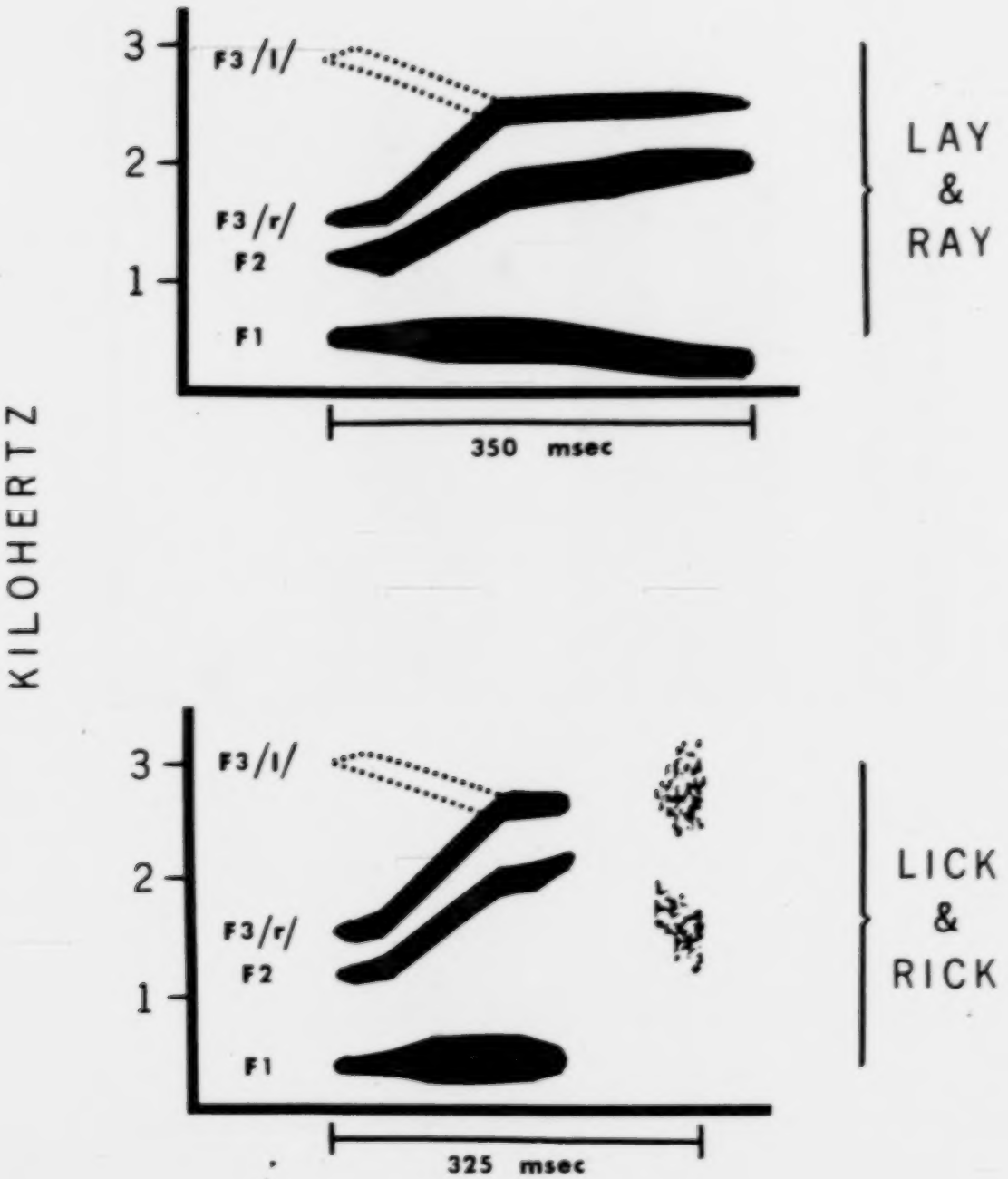


Figure 18: Schematic spectrograms of liquid-initial stimuli and their component parts.

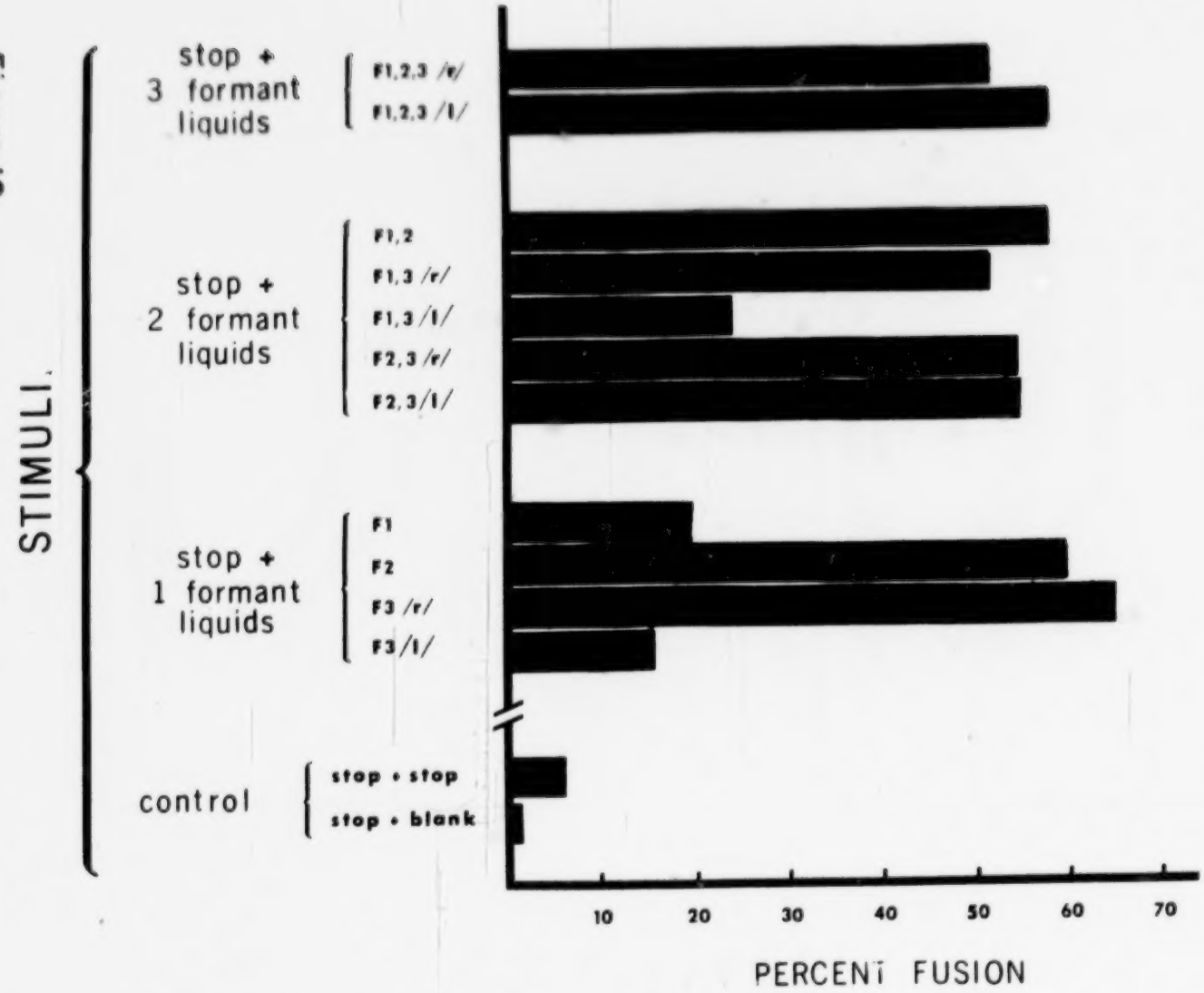


Figure 19

Figure 19: Results of the fusion task for fusible pairs containing various forms of degraded liquids.

Overview. A specific acoustic cue for phonological fusion of stop-liquid stimuli appears to be the liquid formant transition F2 or F3/r/. A single rising formant transition in the range 1000 to 2000 Hz was necessary for fusion to occur.

Experiment IX: Liquid "Chirps"

The previous study found that a mid-range formant transition was necessary for fusion to occur. The present study was designed to determine whether that transition per se is sufficient for fusion when paired with a stop stimulus.

When the F2 transition is removed from a liquid stimulus and synthesized by itself, it sounds similar to a bird's twitter, hence the name "chirp." Mattingly, Liberman, Syrdal, and Halwes (1971) and Wood (1973) have found that these "chirps" are not processed as speech. Chirp stimuli have two general features, relative frequency range and direction (rising vs. falling).

Method. Two stop stimuli were synthesized, PAY and KICK. Liquid stimuli were degraded so that only the F2 transition remained, a 100 msec chirp rising rapidly from a value of 1200 Hz to 1800 Hz. This and other chirps were synthesized at the same amplitude as the F2 transition in the full liquid stimulus. Twelve chirp stimuli were used; there were four frequency values for rising, falling, and steady-state chirps, as shown in Figure 20. Specific stimuli are numbered from low to high representing ordinal position on the frequency scale. Rising chirps were designated with a superscript "r," falling chirps with "f," and steady-state chirps with "s." Endpoints for rising and falling chirps were 600, 1200, 1800, 2400, and 3000 Hz, while steady-state chirps had frequencies of 900, 1500, 2100, and 2700 Hz. The original F2 transition from the liquid stimuli was the chirp 2^r.

In addition to the stop/chirp pairs, there were some control pairs. Ordinary pairs such as PAY/KAY and PAY/LAY were used to obtain baseline fusion rates. Stop/stop pairs were also included to set a lower boundary on fusion rate since Experiment VIII found few stop + liquid responses for such trials. Hence the control pairs provided boundary conditions within which to compare the fusion rates for stop/chirp items. Three lead times were used such that stop/chirp stimulus pairs had the same temporal relationships as the stop and the F2 transition of the full liquid in previous experiments. A dichotic tape of 180 items was prepared: (2 sets of stimuli) x (15 pairs per set)[6] x (3 lead times) x (2 channel arrangements per pair). Twelve subjects listened to one pass through the tape.

Major results. Fusions occurred at a substantially reduced rate for all stop/chirp pairs, as shown in Figure 21. While the fusion rate for stop/liquid control pairs was 47 percent, a rate comparable to previous studies, fusion rates for stop/chirp pairs averaged only 8 percent. The difference was highly significant ($z = 3.18$, $p<.001$).

---

[6] Twelve stop/chirp pairs plus control pairs such as PAY/LAY, PAY/RAY, and PAY/PAY.

Figure 20

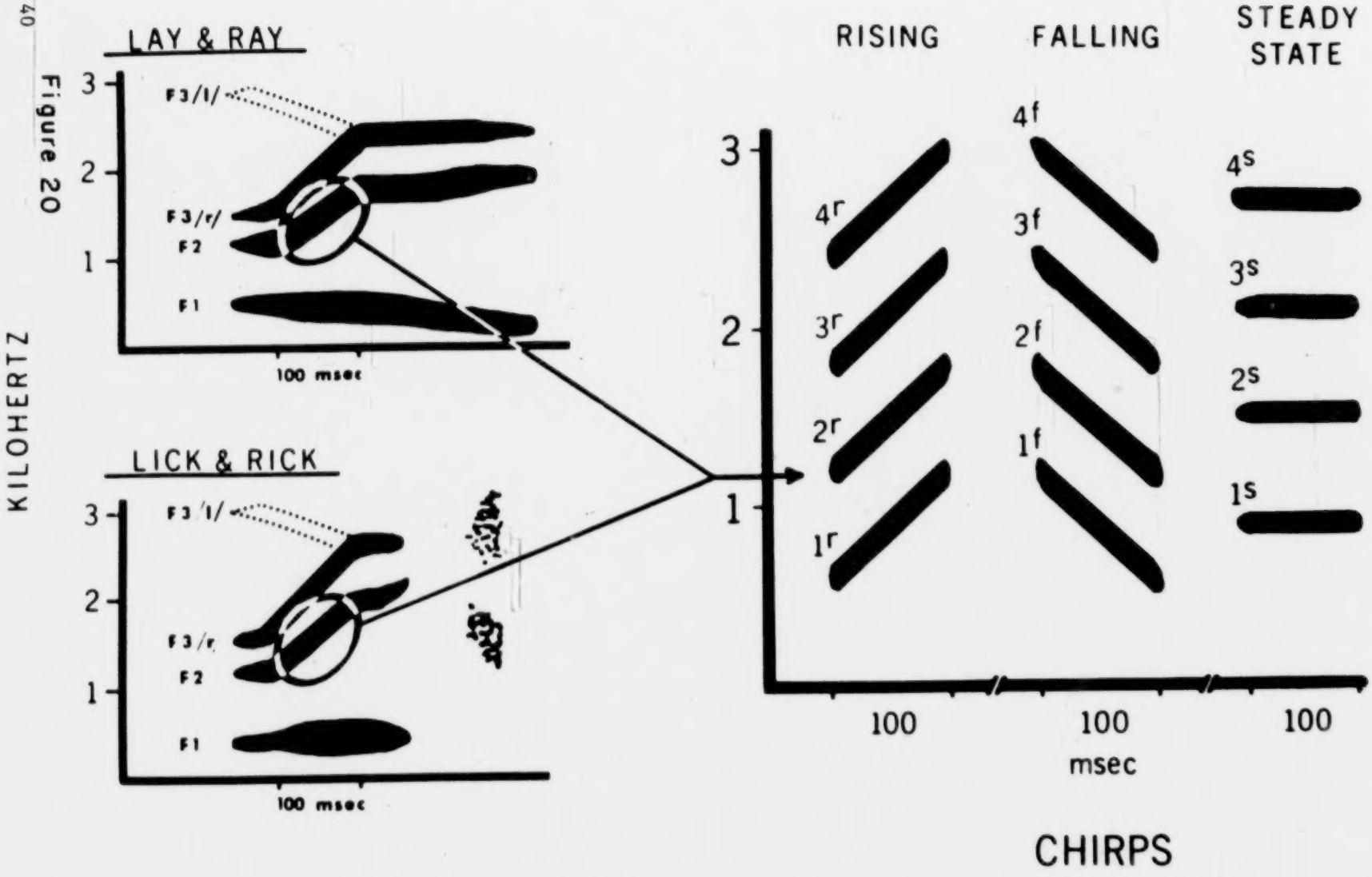


Figure 20: Schematic spectrograms of liquid-initial stimuli and the liquid "chirps."

Figure 21

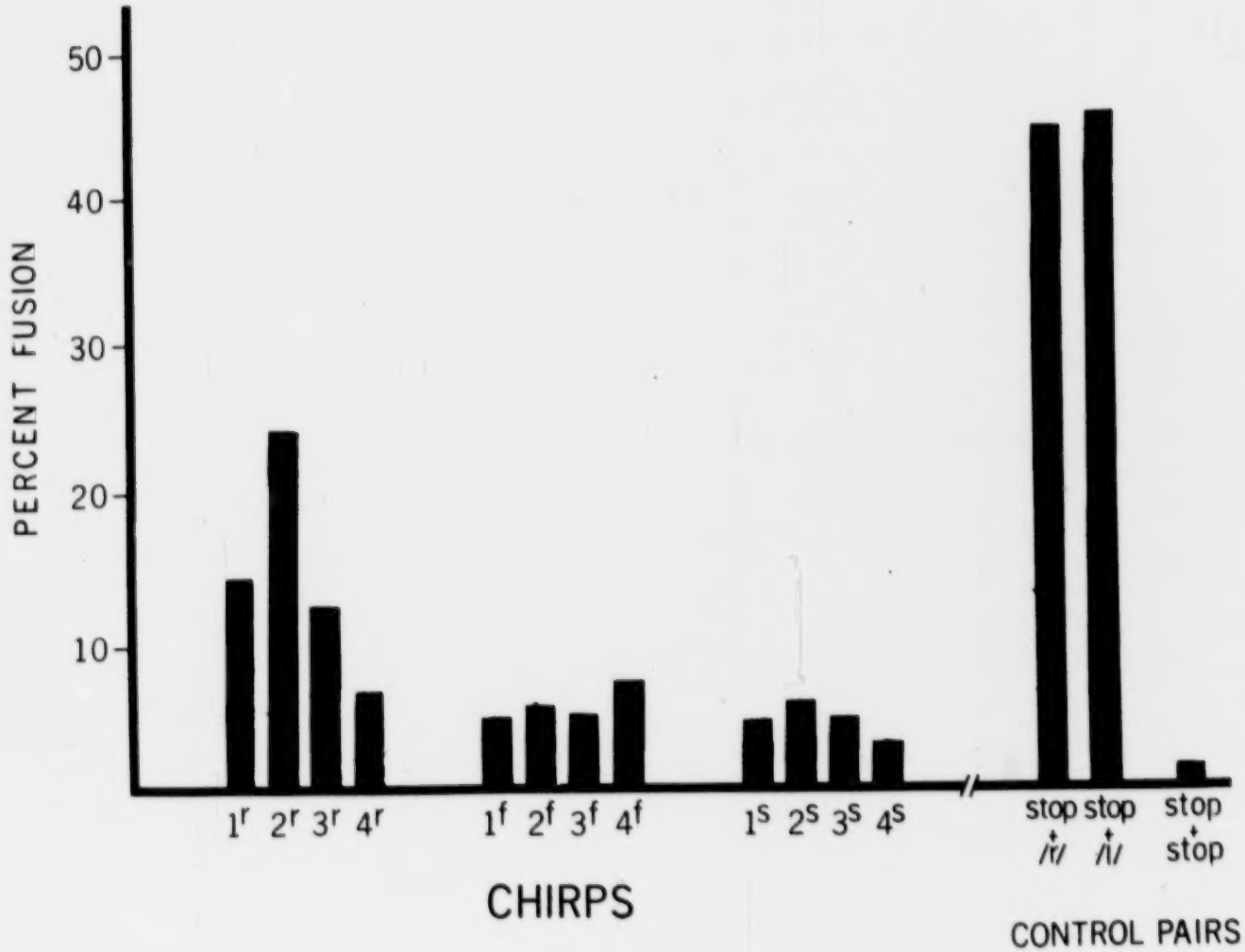


Figure 21: Results of the fusion task for stop/chirp pairs and control pairs.

Fusions did occur, however, for selected stop/chirp pairs. Pairs with rising chirps (1^r, 2^r, 3^r, and 4^r) yielded an average fusion rate of 14 percent, higher than all other stop/chirp pairs combined ($z = 2.6$, $p<.005$). Within the category of rising chirps fusion occurred most readily for stop/2^r pairs where the fusion rate reached 24 percent. Eight of 12 subjects fused at a higher rate for these pairs than for any other stop/chirp pair ($z = 4.0$, $p<.0001$), but even here fusions occurred significantly less often than for the stop/liquid control pairs ($z = 3.18$, $p<.001$). Thus, even the chirp stimulus most appropriate to the full liquid is not entirely sufficient for fusion to occur at an unreduced rate.

Other results. a) The /1/ substitutions occurred for stop/liquid control pairs at rates comparable to previous studies. b) Fusion of stop/chirp pairs, however, were not dominated by stop + /1/ responses. In fact, only stop/2^r pairs yielded more stop + /1/ fusions than stop + /r/, /w/, or /y/ fusions. Fusions for lower frequency chirps (1^r, 1^f, and 1^s) were dominated by stop + /w/ responses, while those for higher frequency chirps (4^r, 4^f, and 4^s) were dominated by stop + /y/ and stop + /1/ responses. c) "False" fusion responses for stop/stop control pairs occurred only 2 percent of the time.

Overview. The fusion rate for stop/chirp pairs was low. Hence the F2 transition in the liquid stimulus was necessary but not entirely sufficient for fusion to occur. Fusions did occur for pairs consisting of only the F2 transition and the stop stimulus, but they occurred much less frequently than for the ordinary stop/liquid pairs. Highest fusion rates among stop/chirp pairs occurred for pairs containing the chirp stimulus whose frequency and direction most nearly matched that of the normal liquid stimulus.

Summary of Linguistic Experiments (IV-IX)

Three levels of linguistic cues were explored (the semantic level, phoneme level, and acoustic level), and the effect of cues at each level on fusion rate was observed. The results suggest that the cognitive processes involved in phonological fusion are influenced by cues at all three linguistic levels. Fusion rate was enhanced when fusible pairs were imbedded in sentence contexts, fusion occurred best for certain classes of phonemes, and specific acoustic cues were found which are important for fusion. Linguistic cues within and outside of the consonant/liquid pairs had a distinct effect on fusion rate.

ADDITIONAL FINDINGS

There are several aspects of the present data which are not primary to the major focus of the paper but which provide additional information about the phenomenon of phonological fusion: a) individual differences in fusion rate, b) changes in fusion responses over time, c) ear effects, and d) /1/ substitutions.

Individual Differences

In order to look at individual differences in fusion rates in the present experiments, subjects were selected from those studies in which the specific experimental variables had little effect on fusion rate. The studies considered

were Experiments II, III, IV, and VI.[7] In addition the results of Cutting (1973a) were also included since the stimuli used in that study were the same as in the present series. Thus, there were 64 subjects in all.

The distribution of fusion rates for these subjects is shown in Figure 22. Note that the distribution is bimodal, with clusters of subjects at the high and low ends. The general shape of this distribution was representative of all experiments in the present series and is similar to that shown by Day (1970a) for subjects who listened to fusible natural speech pairs.

Bimodal individual difference functions in the fusion task appear to reflect different types of processing, and have been found to correlate with performance on other tasks involving the perception of fusible stimuli (Day, 1970a). One such task involves the temporal-order judgment (TOJ) of the initial phonemes in a stop + liquid pair when the stimuli have different relative onset times. Those subjects who fused at a low rate were accurate in judging temporal order. However, those subjects who fused at a relatively high rate were poor judges of temporal order: regardless of whether the stop or liquid stimuli began first in time, they reported that the stop phoneme began first on most trials. Thus, the high fusers were constrained by the phonological rules of English in the TOJ task, and have been called "language-bound" (Day, 1970a). The other group of subjects has been called "stimulus-bound," because they are able to determine accurately the stimulus events. Recent findings suggest that these two groups of subjects retain their group identity for other cognitive abilities such as digit-span memory tasks (Day, 1973), pattern recognition tasks, and secret language tasks (Day, in preparation-c).

Large individual differences appear to occur primarily in higher-level cognitive tasks. Turvey (1973), for example, found that individual differences were greater for higher-level, more central visual processes than for lower-level, more peripheral visual processes. Individual differences have been reported in at least two other higher-order visual tasks. Rommetveit and his co-workers (Rommetveit, Berkeley, and Brøgger, 1968; Rommetveit and Kleven, 1968; Rommetveit, Toch, and Svendsen, 1968a, 1968b) have found marked individual differences in a visual analog to the phonological fusion task. When the letters SHAR are presented to one eye and SHAP to the other, many subjects reported seeing the word SHARP. Other subjects did not fuse the stimuli. Romemtveit called these subjects "nonveridical" and "veridical" perceivers, respectively, and they may be analogous to the language-bound and stimulus-bound subjects in phonological fusion studies. Messmer (cited by Huey, 1908:91ff) also found large individual differences for a dioptic visual task. "Subjective" perceivers were those who, when presented with a stimulus such as INSPECTIXN, never perceived that there was anything amiss. "Objective" perceivers, on the other hand, were quick to report errors in orthography.

Changes in Fusion Responses Over Time

Fusion rates over time were examined for the same 64 subjects discussed in the individual differences section. The top half of Figure 23 shows fusion rate divided in quartiles of the test. Fusion rate was stable throughout the tests, averaging about 60 percent for each quartile across the various experiments.

---

[7] For Experiment IV only the no-sentence context condition was considered.

Figure 22

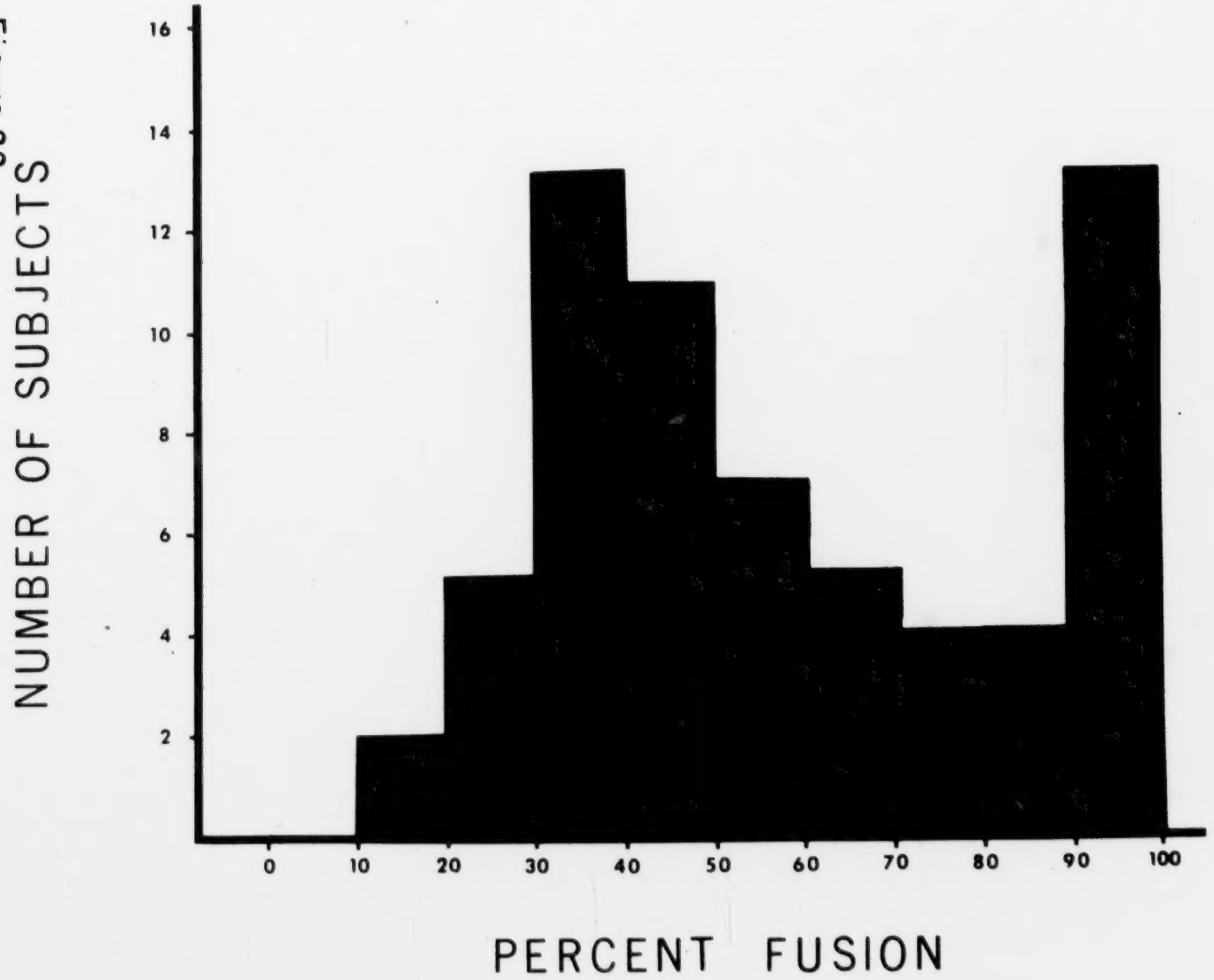


Figure 22: Distribution of fusion rates for 64 subjects.

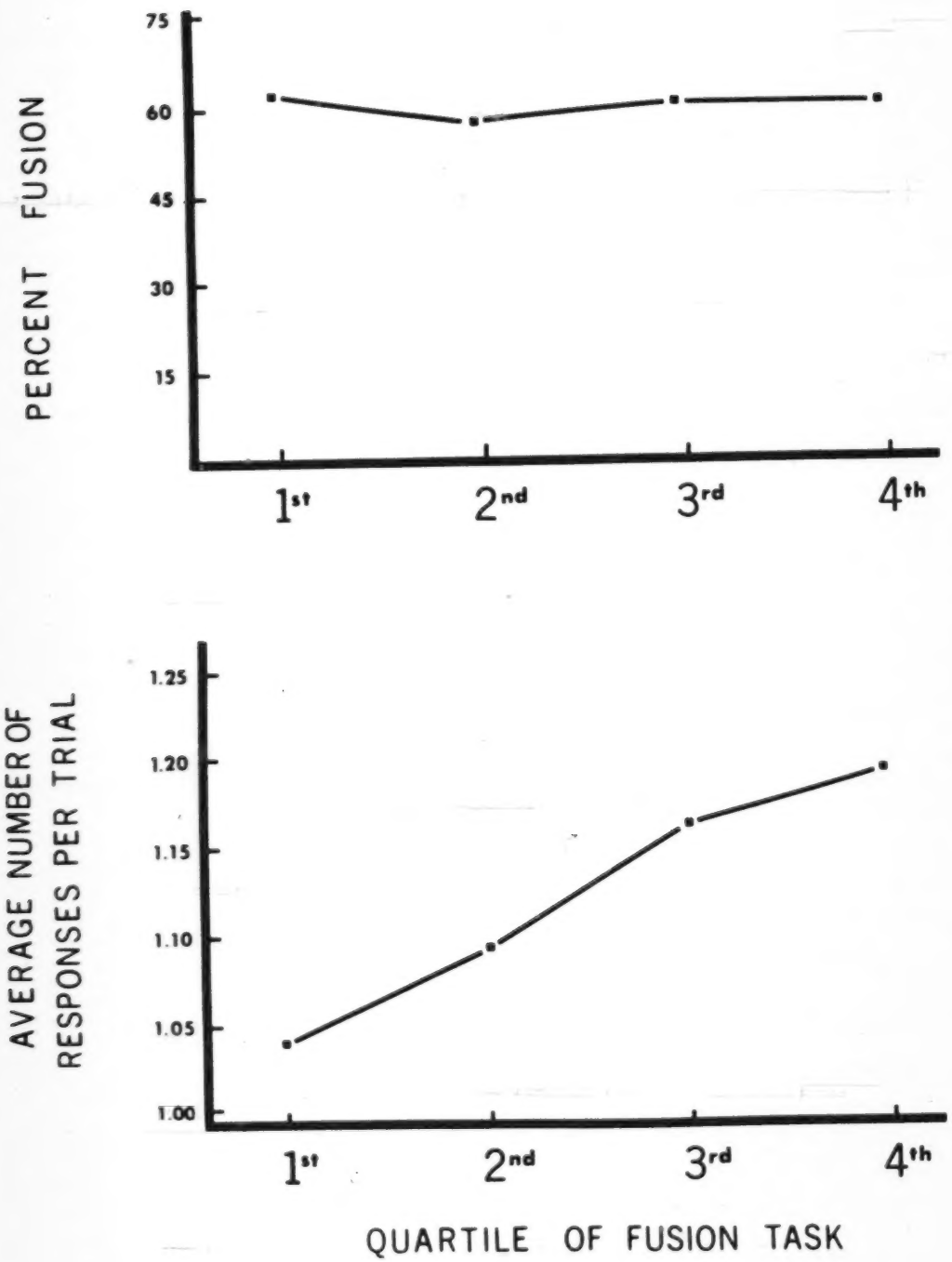


Figure 23: Percent fusion and average number of responses for each quarter of the fusion tasks.

Thus, the subjects appeared to be no more "informed" about the nature of the stimuli on the last trial than they were on the first. This conclusion, however, must be qualified somewhat since fusion rate by itself does not reflect the entire nature of the subjects' responses.

During the course of the task, subjects began to report hearing more than one item per trial. Given the stimuli PAY/LAY, most responses were single items, for example, PLAY or PAY. However, double responses occurred with increasing frequency over the course of the task. For example, subjects reported PAY and LAY, PLAY and PAY, PLAY and LAY, or even PLAY and PLAY. As shown at the bottom of Figure 23 double responses occurred on only 1 in 25 trials in the first quartile, and increased linearly to 1 in 5 trials by the last quartile. This increase indicates that subjects became more aware that two different stimuli were presented on each trial as the task progressed. However, since fusion rates remained constant throughout the task, it is clear that the subjects were still unaware of the specific nature of the separate stimuli.

Ear Effects

Fusion occurred equally often for cases where the stop-initial stimulus was presented to the right ear and the liquid-initial stimulus to the left ear, and for the reverse configuration. Hence there were no ear effects for fused trials. Many experimenters have found a right-ear advantage for dichotic speech stimuli (for a review, see Studdert-Kennedy and Shankweiler, 1970). Previous phonological fusion studies, however, did not yield ear differences, nor were any obtained in the present experiments. Ear differences for speech stimuli occur only when the items cannot be combined into a single percept. In such cases an input competes with its opposite-ear rival for a processor which typically cannot handle both of them at the same time. Information loss results from this competition and ear effects reflect the general loss of information from each ear. In phonological fusion there is no information loss: if the stimuli are PAY/LAY and the subject reports hearing PLAY, he has reported all of the linguistic units presented to him, reorganized into a perceptual whole. Without the loss of information there can be no decrement in overall performance, and hence no ear effect can result.

Ear advantages also failed to occur for nonfused trials. For single responses the right-ear stimulus was reported 50 percent of the time and the left-ear stimulus 50 percent of the time. Typically in double responses all of the information in both stimuli was reported, and therefore no ear effect could result. Ear scores can also be measured in terms of which item was reported first, the right-ear stimulus or the left-ear stimulus, but in the present series of studies this analysis again yielded no ear effects.

The /l/ Substitutions

Given the stimuli PAY/RAY subjects often reported hearing PLAY. Difficulties with /r/ vs. /l/ occur in a wide variety of other situations besides phonological fusion. Children, for example, have more difficulty in pronouncing /r/ than /l/ and sometimes pronounce both phonemes as /l/ (Morley, 1957; Powers, 1957; Murray, 1962); the deaf have more trouble processing /r/ than /l/, and often hear /l/ in both cases (Rosen, 1962); /r/ has a less stable articulation pattern than /l/ (Bronstein, 1960; Delattre, 1967); /r/ may yield more metathesis [spoonerism] errors than /l/ (see Cutting and Day, 1972); /r/ is more

difficult to pronounce correctly under conditions of delayed auditory feedback than /l/ (Applegate, 1968); and /r/ yields a more varied pattern of ear advantages than /l/ in certain dichotic listening tasks not involving fusion (Cutting, 1973d). Stress on perception and production systems, then, is more harmful to /r/ than /l/. The dichotic fusion results are complementary to these studies.

In phonological fusion /l/ substitutions cannot be accounted for by stimulus identifiability. In fact, /r/ stimuli were slightly better identified than /l/ stimuli (see Appendix B). Word frequency and cluster frequency cannot account for the substitutions either. Furthermore, the /l/-substitution effect appears to override all linguistic cues except the semantic cues at the word level.

Cutting (1973b) found that subjects not only reported hearing PLAY when PAY/LAY and PAY/RAY were presented, for example, but that they also could not discriminate between the fusible pairs. In other words, PAY/LAY and PAY/RAY not only tended to yield the same fusion response, but they also were virtually indistinguishable. These results suggest that there may be specific acoustic cues in the dichotic listening situation which might be pertinent for the perception of stop + /l/ clusters and improper for the perception of stop + /r/.

GENERAL DISCUSSION

The studies in the present paper showed that phonological fusion was vigorously independent of nonlinguistic stimulus variation, but sensitive to linguistic variation at all the levels that were studied. This overview suggests that there is a marked difference in the way linguistic and nonlinguistic stimulus dimensions may be processed.

Higher- and lower-level dimensions. Linguistic and nonlinguistic dimensions of auditory stimuli are hierarchically related: linguistic dimensions imply the existence of nonlinguistic dimensions, whereas nonlinguistic dimensions may occur without any linguistic dimension present. For example, it is impossible to have a stimulus (such as the word BEE) which has linguistic attributes but no nonlinguistic attributes, such as time constraints, pitch, and intensity. On the other hand, it is commonplace to have a stimulus (such as a tone) which has nonlinguistic attributes but has no linguistic properties. These attributes have been called higher-level and lower-level stimulus dimensions.

Higher- and lower-level processing. Given their hierarchical relationship one might assume that the processing of nonlinguistic dimensions necessarily occurs before linguistic processing, and that linguistic analyses might be contingent on the outcome of previous nonlinguistic analyses. Indeed, for some tasks this appears to be the case. Day and Wood (1972) and Wood (1973), for example, found results supporting this notion in diotic forced-choice reaction time tasks: irrelevant variation along a nonlinguistic dimension impeded performance on a task involving linguistic decisions, whereas irrelevant linguistic variation had little effect on nonlinguistic task performance.

Dichotic phonological fusion, however, demonstrates that linguistic analysis can occur independent of nonlinguistic constraints on the to-be-fused stimuli. Fusible stimuli can vary widely in relative onset, pitch, and intensity with

little if any effect on perception. When different stimuli are presented to opposite ears, the interaction that occurs between them is qualitatively different from that which occurs for stimuli presented diotically or monaurally (Studdert-Kennedy, Shankweiler, and Schulman, 1970; Brady-Wood and Shankweiler, 1973; Cutting, 1973e). In the diotic and monaural cases, the interaction occurs at a lower level, where timing, pitch, and intensity are important. In the dichotic case the constraints of these dimensions may be by-passed so that stimuli may interact at a higher level.

However, not all dichotic tasks are free from nonlinguistic bonds. For example, three of the six auditory fusions mentioned at the beginning of this paper (and discussed in more detail in Cutting, 1972) are sensitive to timing, pitch, and intensity differences between the stimuli. These are lower-level fusions which occur for both speech and nonspeech stimuli. The higher-level fusions, which are not dependent on nonlinguistic cues in the stimuli, occur only for speech sounds. In such cases, it is not raw stimuli that interact in the higher-level processor, but linguistically coded information in the form of phonemes or phonetic features.

The process of phonological fusion. Higher-level fusions appear to occur solely in the language processor where fusible linguistic units may be perceptually combined. The nonlinguistic dimensions of the stimuli have either been discarded at this point or sent to a different processor. Yet when the fusible stimuli entered the system they carried both linguistic and nonlinguistic attributes on the same waveforms. The process model considered previously (Figure 7) is helpful in explaining how linguistic and nonlinguistic information might be separated, and how phonological fusion occurs.

The higher-level processor is an information coder which codes incoming speech signals at enormous savings in terms of the amount of information needed to be stored. Liberman, Mattingly, and Turvey (1972) have estimated that this coder typically reduces a 40,000 bit-per-second acoustic signal into a 40 bit-per-second phonetic code suitable for further linguistic analysis. The cost of this coding process, however, is the quick loss in availability of nonlinguistic information. The linguistic attributes of the signal are digitized (coded), while the nonlinguistic dimensions appear to remain in nondigital, raw form.

Dichotic stimuli appear to maintain their separate integrities while being transmitted primarily to the hemisphere opposite from the ear of arrival (Milner, Taylor, and Sperry, 1968; Sparks and Geschwind, 1968). While language processing occurs primarily in the left hemisphere for most people, a certain amount of linguistic coding may occur in the right hemisphere (see Gazzaniga, 1967). Thus two speech signals, one from the right ear and one from the left, may be analyzed (coded) independently in separate hemispheres. Phonological fusion, then, appears to be the integration of two coded representations of the fusible stimuli.

Semantic, phonemic, and acoustic dimensions of the fusible stimuli often have an effect on fusion rate. It appears that once the fusible stimuli have been linguistically coded, this coded information interacts with other linguistic information at different levels of language. Given the appropriate experimental situation, cues from all three levels appear to work in concert. Consider one of the sentence pairs from Experiment IV: THE TREES ARE GOING AGAIN

presented to one ear and THE TREES ARE ROWING AGAIN presented to the other ear. Semantic cues at the sentence level increased the fusion rate for GO/ROW pairs beyond the rate they yielded when presented in isolation. Cues at the phoneme level were also influential in maintaining a high fusion rate. For example, Experiment V found that GO/ROW pairs fused at a higher rate than FOE/ROW pairs, indicating that certain phoneme classes fuse more readily than others. Experiment VIII found that the second formant transition in the liquid was a specific cue for fusion, and when it was not present the fusion rate was considerably reduced. Furthermore, this cue appears to be pertinent to the /l/-substitution effect: fusion rates and /l/ substitutions were approximately equal for stop/F2 and stop/liquid stimuli. Thus the high rate of THE TREES ARE GLOWING AGAIN responses appears to be the result of the synergy of cues from three very different linguistic levels.

APPENDIX A - ACOUSTIC STRUCTURE OF STIMULI

Stimuli within a particular set were identical in all respects except for the acoustic structure of the first 150 msec. Stop stimuli began with appropriate 50 msec transitions ending at the steady-state frequencies of the following vowel. Liquid stimuli followed a pattern suggested by O'Connor et al. (1957): each liquid began with a 50 msec steady-state onglide, followed by 100 msec transition in F2 and F3 and a 20 msec transition in F1, followed by a vowel. Within a particular set, liquid stimuli differed only in the steady-state onglide of F3, and in the F3 transition, the cue most important for the separation of /r/ from /l/ (O'Connor et al., 1957).

Since previous studies had found many /l/ substitutions, an effort was made to make /r/ stimuli as highly identifiable as possible. Thus, when there was any doubt as to what values were to be selected for /r/ and /l/ stimuli, decisions were always made to favor /r/ rather than /l/. The choice in duration of the steady-state onglides, the duration of the F2 and F3 transitions, and the frequency value of the F1 steady-state onglide (311 Hz) were three such decisions. Identification results proved that /r/ stimuli were somewhat more identifiable than /l/ stimuli (see Appendix B). Thus, /l/ substitutions cannot be accounted for by the identifiability of the /r/ stimuli.

APPENDIX B - IDENTIFICATION TASK RESULTS

Subjects participated in identification tasks after fusion tasks so that specific information about the individual stimuli gained in the identification task could not influence their fusion results. Ten tokens of each stimulus used in a particular experiment were presented singly in a random order with three seconds between items. Subjects were instructed to write down what they heard after each presentation. In some experiments they were free to write whatever they heard, while in others they chose among a limited repertoire of initial phonemes. The results were the same regardless of the instructions: averaging over all experiments, stop stimuli were correctly identified on 95 percent of all trials, fricative stimuli 96 percent, /r/ stimuli 94 percent, /l/ stimuli 90 percent, and semivowel stimuli (/w/ and /y/) 83 percent. When errors occurred they were primarily within-phoneme-class errors. Thus, stops were identified as stops, fricatives as fricatives, and liquids and semivowels as liquids and semivowels. Identification results were similar for all subjects in all experiments.

APPENDIX C - SEMANTIC APPROPRIATENESS OF SENTENCES

Twenty subjects who did not participate in any of the fusion experiments were asked to rate which of two alternative forms of a sentence was the most "meaningful": for example, THE MINISTER PRAYS FOR US vs. THE MINISTER PLAYS FOR US. All sentences were possible fusion responses in Experiment IV. Results shown in Table C-1 indicated that most subjects agreed as to which sentence of each pair was most "meaningful" and these ratings were taken as a measure of the semantic appropriateness of the sentences.

TABLE C-1: Forced-choice scores for the possible fused sentences.

Sentence Pair	No. Subjects	
a. THE MINISTER PRAYS FOR US*	18	$z = 3.8$, $p<.0001$
b. THE MINISTER PLAYS FOR US	2	
a. THE TRUMPETER PRAYS FOR US	1	$z = 4.0$, $p<.0001$
b. THE TRUMPETER PLAYS FOR US*	19	
a. THE TREES ARE GROWING AGAIN*	18	$z = 3.8$, $p<.0001$
b. THE TREES ARE GLOWING AGAIN	2	
a. THE COALS ARE GROWING AGAIN	1	$z = 4.0$, $p<.0001$
b. THE COALS ARE GLOWING AGAIN*	19	

*Semantically appropriate sentences.

REFERENCES

Applegate, R. B. (1968) Segmental analysis of articulatory errors under delayed auditory feedback. Project on Linguistic Analysis, University of California at Berkeley 8, 1-27.

Brady-Wood, S. and D. Shankweiler. (1973) Effects of attenuation of one of two channels on perception of opposing pairs of nonsense syllables when monotically and dichotically presented. Paper presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., April.

Broadbent, D. E. (1955) A note on binaural fusion. Quart. J. Exp. Psychol. 7, 46-47.

Broadbent, D. E. and P. Ladefoged. (1957) On the fusion of sounds reaching different sense organs. J. Acoust. Soc. Amer. 29, 708-710.

Bronstein, A. J. (1960) The Pronunciation of American English. (New York: Harper and Row) 111ff.

Carroll, J. B., P. Davies, and B. Richman, eds. (1971) Word Frequency Book. (New York: Houghton and Mifflin).

Cooper, F. S. and I. G. Mattingly. (1969) Computer-controlled PCM system for investigation of dichotic speech perception. J. Acoust. Soc. Amer. 46, 115(A).

Cutting, J. E. (1972) A preliminary report on six fusion in auditory research. Haskins Laboratories Status Report on Speech Research SR-31/32, 93-107.

Cutting, J. E. (1973a) Phonological fusion in synthetic and natural speech. Haskins Laboratories Status Report on Speech Research SR-33, 19-27.

Cutting J. E. (1973b) Speech misperception: Inferences about a cue for cluster perception from a phonological fusion task. Haskins Laboratories Status Report on Speech Research SR-33, 51-65.

Cutting, J. E. (1973c) Perception of speech and nonspeech, with and without transitions. Haskins Laboratories Status Report on Speech Research SR-33, 37-46.

Cutting, J. E. (1973d) A parallel between degree of encodedness and the ear advantage: Evidence from an ear-monitoring task. J. Acoust. Soc. Amer. 53, 358(A). (Also in Haskins Laboratories Status Report on Speech Research SR-29/30 as: A parallel between encodedness and the magnitude of the right-ear effect.)

Cutting, J. E. (1973e) Phonological fusion of synthetic stimuli in dichotic and binaural presentation modes. Haskins Laboratories Status Report on Speech Research SR-34 (this issue).

Cutting, J. E. (1973f) Phonological fusion of stimuli produced by different vocal tracts. Haskins Laboratories Status Report on Speech Research SR-34 (this issue).

Cutting, J. E. and R. S. Day. (1972) Dichotic fusion along an acoustic continuum. J. Acoust. Soc. Amer. 52, 175(A).

Cutting, J. E. and R. S. Day. (in preparation) Multidimensional stimulus variation in a phonological fusion task.

Day, R. S. (1968) Fusion in dichotic listening. Unpublished Ph.D. thesis, Stanford University (Psychology).

Day, R. S. (1970a) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Haskins Laboratories Status Report on Speech Research SR-21/22, 71-87.

Day, R. S. (1970b) Temporal order perception of a reversible phoneme cluster. Haskins Laboratories Status Report on Speech Research SR-24, 47-56.

Day, R. S. (1973) Digit span memory in language-bound and stimulus-bound subjects. Paper presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., April.

Day, R. S. (in preparation-a) Fusion despite knowledge of the stimuli.

Day, R. S. (in preparation-b) Release from language-bound perception.

Day, R. S. (in preparation-c) Individual differences in cognition.

Day, R. S. and J. E. Cutting. (1970) Levels of processing in speech perception. Paper presented at the 10th annual meeting of the Psychonomic Society, San Antonio, Texas, November.

Day, R. S. and J. E. Cutting. (1971) What constitutes perceptual competition in dichotic listening? Paper presented at the annual meeting of the Eastern Psychological Association, New York, April.

Day, R. S., J. E. Cutting, and P. M. Copeland. (1971) Perception of linguistic and nonlinguistic dimensions of dichotic stimuli. Paper presented at the 11th annual meeting of the Psychonomic Society, St. Louis, Mo., November. (Also in Haskins Laboratories Status Report on Speech Research SR-27, 193-198.)

Day, R. S. and C. C. Wood. (1972) Interaction between linguistic and nonlinguistic processing. J. Acoust. Soc. Amer. 51, 79(A). (Also in Haskins Laboratories Status Report on Speech Research SR-27, 185-192.)

Delattre, P. C. (1967) A dialect study of American /r/ by X-ray motion picture. The General Phonetic Characteristics of Languages, Final Report, University of California at Santa Barbara, 7-80.

Denes, P. B. (1965) On the statistics of spoken English. J. Acoust. Soc. Amer. 35, 892-904.

Gazzaniga, M. (1967) The split brain in man. Sci. Amer. 217, 24-29.

Halwes, T. G. (1969) The effects of dichotic fusion on the perception of speech. Ph.D. thesis, University of Minnesota (Psychology). (Issued as Supplement to Haskins Laboratories Status Report on Speech Research.)

Huey, E. B. (1908) The Psychology and Pedagogy of Reading. (Republished by MIT Press, Cambridge, 1968.)

Hultzén, L., J. Allen, and M. Miron. (1964) Tables of Transitional Frequencies of English Phonemes. (Urbana, Ill.: University of Illinois Press).

Lehiste, I. (1962) Acoustical characteristics of selected English consonants. University of Michigan Communication Sciences Laboratory Report No. 9, 1-167.

Liberman, A. M. (1957) Some results of research on speech perception. J. Acoust. Soc. Amer. 29, 117-123.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Liberman, A. M., F. Ingemann, L. Lisker, P. C. Delattre, and F. S. Cooper. (1959) Minimal rules for synthesizing speech. J. Acoust. Soc. Amer. 31, 1490-1499.

Liberman, A. M., I. G. Mattingly, and M. T. Turvey. (1972) Language codes and memory codes. In Human Memory, ed. by A. W. Melton and E. Martin. (Washington, D. C.: V. H. Winston) 307-333.

Licklider, J. C. R. (1951) Basic correlates of the auditory stimulus. In Handbook of Experimental Psychology, ed. by S. S. Stevens. (New York: John Wiley) 985-1039.

Lisker, L. (1957) Minimal cues for separating /w, r, l, y/ in intervocalic position. Word 13, 257-267.

Mattingly, I. G. (1968) Synthesis by rule of General American English. Ph.D. thesis, Yale University (English). (Issued as Supplement to Haskins Laboratories Status Report on Speech Research.)

Mattingly, I. G., A. M. Liberman, A. K. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. 2, 131-157.

Milner, B., L. Taylor, and R. W. Sperry. (1968) Lateralized suppression of dichotically presented digits after commissure section in man. Science 161, 184-185.

Morley, M. E. (1957) Development of Speech Disorders in Childhood. (London: Livingstone).

Murray, R. S. (1962) The development of /r/ in the speech of preschool children. Unpublished Ph.D. thesis, Stanford University (Speech).

O'Connor, J. D., L. J. Gerstman, A. M. Liberman, P. C. Delattre, and F. S. Cooper. (1957) Acoustic cues for the perception of initial /w, y, r, l/ in English. Word 13, 25-43.

Pisoni, D. B. (1971) On the nature of categorical perception of speech sounds. Ph.D. thesis, University of Michigan (Psycholinguistics). (Issued as Supplement to Haskins Laboratories Status Report on Speech Research.)

Powers, M. H. (1957) Functional disorders of articulation: Symptomatology and etiology. In Handbook of Speech Pathology, ed. by L. E. Travis. (New York: Appleton-Century-Crofts) 707-768.

Rommetveit, R., M. Berkeley, and J. Brøgger. (1968) Generation of words from tachistoscopically presented nonword strings of letters. Scandinavian J. Psychol. 9, 150-156.

Rommetveit, R. and J. Kleven. (1968) Word generation: A replication. Scandinavian J. Psychol. 9, 277-281.

Rommetveit, R., H. Toch, and D. Svendsen. (1968a) Effects of contingency and contrast on the cognition of words. Scandinavian J. Psychol. 9, 138-144.

Rommetveit, R., H. Toch, and D. Svendsen. (1968b) Semantic, syntactic, and associative context effects in stereoscopic rivalry situation. Scandinavian J. Psychol. 9, 145-149.

Rosen, J. (1962) Phoneme identification in sensorineural deafness. Unpublished Ph.D. thesis, Stanford University (Speech).

Sparks, R. and N. Geschwind. (1968) Dichotic listening in man after section of neocortical commissures. Cortex 4, 3-16.

Studdert-Kennedy, M. and D. P. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

Studdert-Kennedy, M., D. P. Shankweiler, and S. Schulman. (1970) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.

Thorndike, E. L. and I. Lorge. (1944) The Teacher's Word Book of 30,000 Words. (New York: Teachers College Press).

Turvey, M. T. (1973) On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. Psychol. Rev. 80, 1-52.

Wood, C. C. (1973) Levels of processing in speech perception: Neurophysiological and information-processing analyses. Unpublished Ph.D. thesis, Yale University (Psychology).

Wood, C. C., W. R. Goff, and R. S. Day. (1971) Auditory evoked potentials during speech perception. Science 173, 1248-1251.

Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation Modes

James E. Cutting[+]
Haskins Laboratories, New Haven, Conn.

Phonological fusion occurs when the phonemes of two different speech stimuli are combined into a single percept which contains all the linguistic information from the two inputs. Thus, for example, PAHDUCT/RAHDUCT⟶PRODUCT (Day, 1968) and BANKET/LANKET⟶BLANKET (Day, 1970).[1] The present study was designed to observe fusion rates in both dichotic and binaural presentation modes, shown in Figure 1. Previous studies (for example, Day, 1968, 1970; Cutting, 1973a) have presented fusible stimuli dichotically, one stimulus to the right ear and the other to the left. Such presentation appears to allow for the independent processing of the two items before combining them into a perceptual whole (Cutting, 1973a). In the binaural mode, on the other hand, the independent extraction of linguistic features from the two stimuli is not possible since the two stimuli are electrically mixed and both are presented to each ear as part of the same waveform.[2]

Method. Four stimulus sets of the same general pattern were synthesized on the Haskins Laboratories' parallel resonance synthesizer: the PAY set (PAY, RAY, LAY), the BED set (BED, RED, LED), the CAM set (CAM, RAM, LAMB), and the GO set (GO, ROW, LOW). All sets had been used previously by Cutting and Day (1972) and Cutting (1973b, 1973c). Stimuli within a given set were identical in duration, pitch, and intensity, and differed only in the formant transition requisites of the first 150 msec. Each stimulus was highly identifiable when presented in isolation (Cutting, 1973a, 1973b). Fusible pairs were constructed using stimuli within the same set; for example, PAY/RAY and PAY/LAY. All stimuli and possible fusions were high frequency English words (Carroll, Davies, and Richman, 1971). Stimuli were digitized and stored on disc file for the preparation of dichotic and binaural fusible pairs. Pairs of stimuli began simultaneously, or one

---

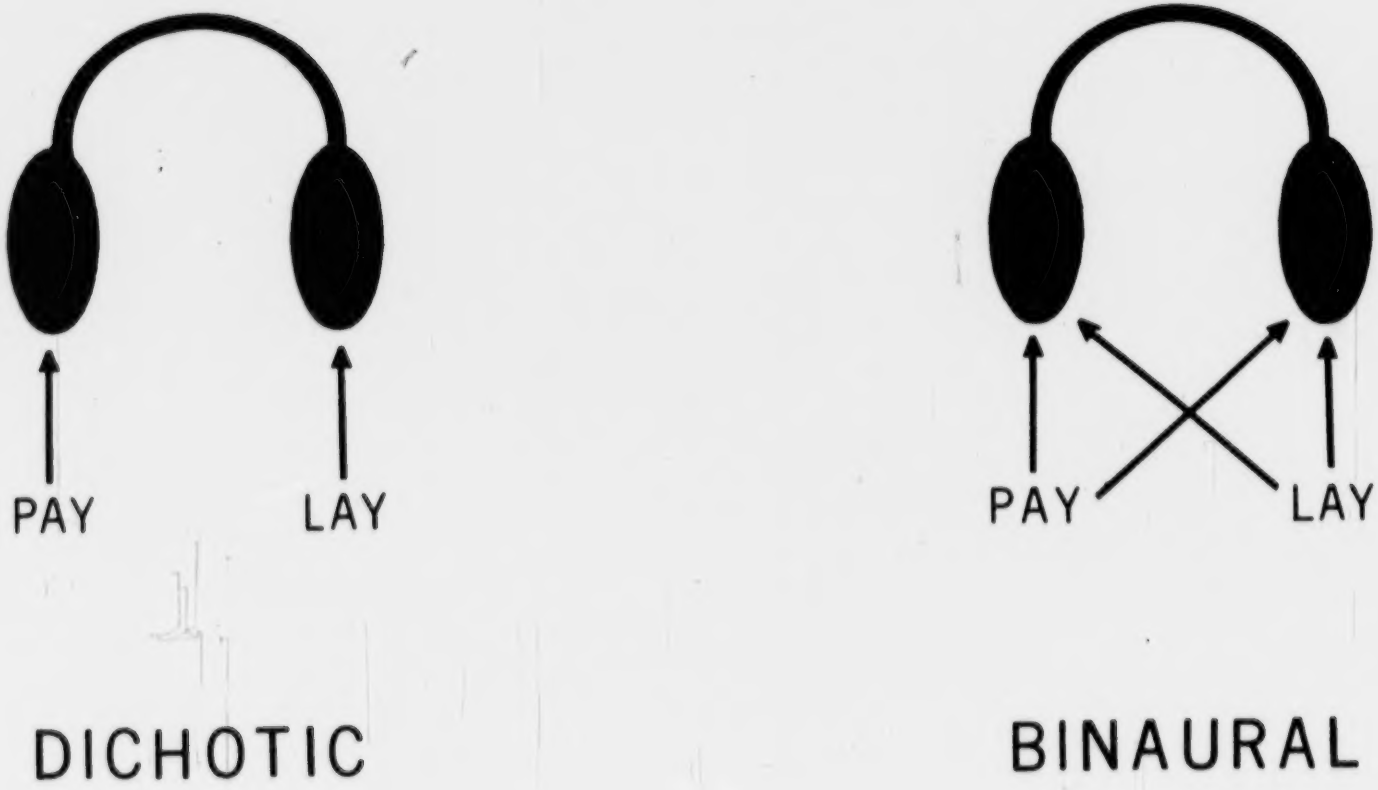
[+]Also Yale University, New Haven, Conn.

[1]The arrow should be read as "yields."

[2]Licklider (1951:1027) noted that binaural presentation is a general term denoting the stimulation of both ears, while diotic presentation is a specific term for the presentation of a single stimulus to both ears at the same time. Since, in the present study, two stimuli are both presented to both ears, the more general term binaural is used.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

Figure 1



MODES OF PRESENTATION

Figure 1: The presentation modes of the fusible pair PAY/LAY.

stimulus preceded the other by 50 msec. LAY, for example, began before PAY, and PAY before LAY on an equal number of trials. Channel assignments were counterbalanced for each dichotic pair.

Twenty-four Yale University undergraduates listened to both dichotic and binaural fusible items. Eight subjects listened first to a sequence of dichotic trials, then to a sequence of binaural trials (Group 1), while eight others listened in the reverse order (Group 2). The remaining subjects listened to a tape consisting of dichotic and binaural trials randomly intermixed (Group 3).

Two tapes were prepared. The first tape consisted of 96 dichotically recorded items: (4 sets of stimuli) x (2 stop/liquid pairs per set) x (3 lead times) x (2 channel arrangements) x (2 observations per pair). The tape was played on an Ampex AG-500 dual-track tape recorder, sent through a mixing box and a listening station to Grason Stadler earphones (Model TDH39-300Z). At the mixing box signals either remained separate (dichotic) or were mixed onto both channels (binaural) according to the experimental condition. The second tape was exactly twice as long (192 items) and consisted of a random sequence of all possible dichotic and binaural pairs. Binaural items on this tape were constructed as the tape was recorded. Subjects wrote down what they heard on each trial.

Results. Fusion rates were much higher for dichotic pairs than for binaural pairs--45 percent and 15 percent, respectively. This 3:1 ratio was highly significant ($z = 4.8$, $p<.0001$): all 24 subjects yielded fusion results in this direction, and as shown in Figure 2, each group of subjects yielded fusion rates indicative of this difference.

Fusion rates were highest for dichotic pairs in Group 1, when they were presented in a blocked manner before the binaural trials, and lowest for Group 2, when they were presented after the binaural block of trials. There was, however, no significant difference among the three groups of subjects. For binaural pairs, fusion rates were highest for Group 3 in the mixed presentation, and lowest for Group 2 in the blocked presentation when they preceded the dichotic trials. The difference between these two groups was significant [U (8,8) = 4, $p<.001$], indicating that fusion rates for binaural pairs was increased when they were presented in the same random sequence with dichotic fusible pairs, while fusion rate for dichotic pairs remained relatively stable.

Stop + /r/ stimuli (for example, PAY/RAY) and stop + /l/ stimuli (PAY/LAY) fused equally well. The fusion responses, however, showed that most fusions were stop + /l/. In the fusion response, /l/ was often substituted for /r/ (PAY/RAY⟶PLAY), whereas the reverse substitution rarely occurred. Day (1968), Cutting and Day (1972), and Cutting (1973a, 1973b, 1973c) have observed the /l/-substitution effect. It cannot be accounted for by word frequency of the fusion responses or cluster frequency of stop consonants and liquids, but it may be linked to the specific cues in the dichotic situation (Cutting, 1973c). The /l/ substitutions occurred to a lesser degree in the binaural situation.

The effect of lead time was similar to that found in previous studies using these stimuli (Cutting, 1973b, 1973c): fusions occurred more readily when the stop stimulus began before the liquid than in either of the other two lead-time conditions. This effect occurred for both dichotic and binaural items. For

Figure 2

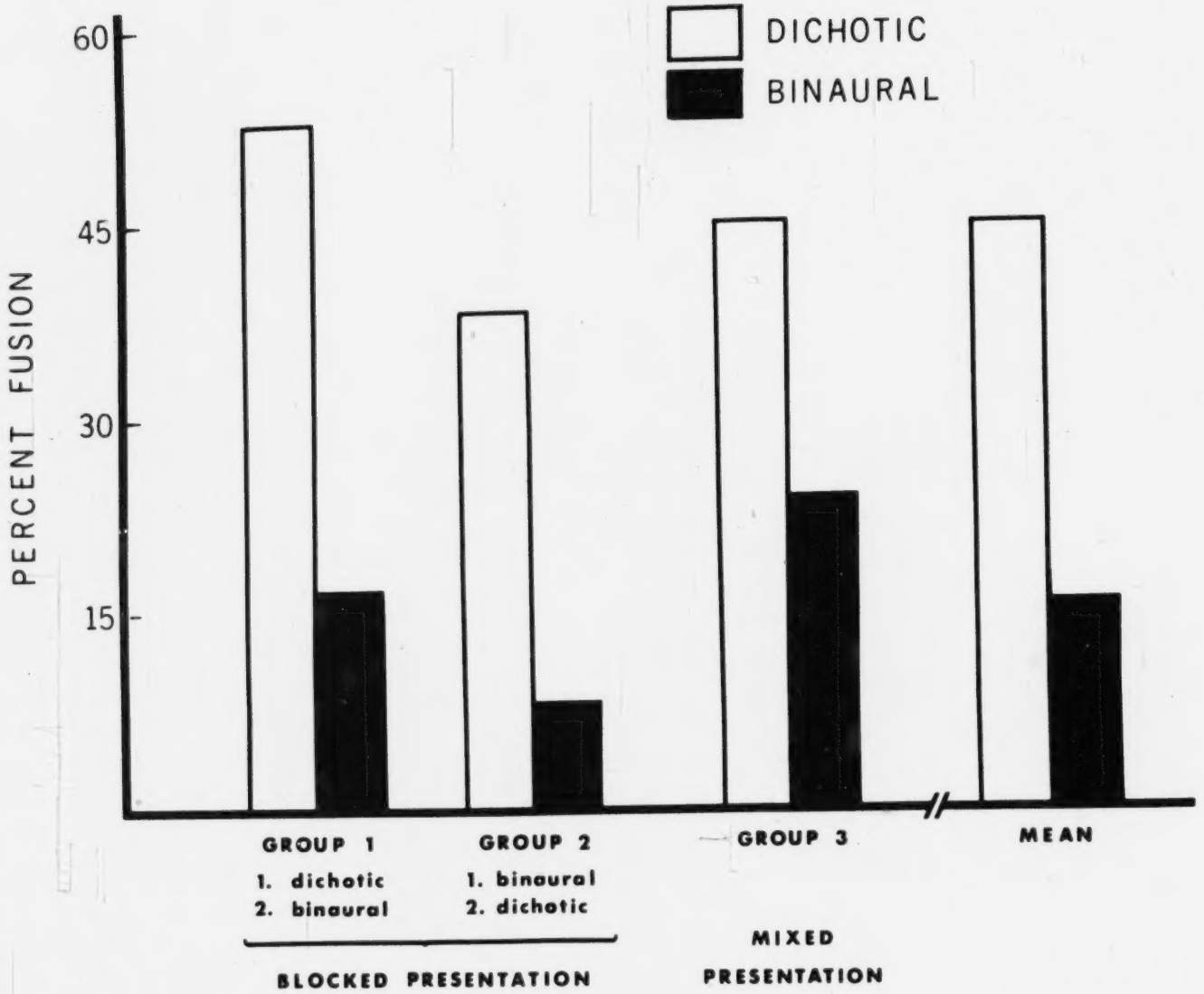


Figure 2: Fusion rates of dichotic and binaural items for three experimental groups.

further discussion of the effects of lead time see Cutting (1973a) and Day (in preparation).

In both conditions, when fusion responses did not occur, typically only the stop stimulus was reported--for example, PAY/LAY⟶PAY.

Conclusion. The difference between dichotic and binaural fusion rates may be interpreted as an indication that phonological fusion is a higher-level process. Higher-level fusions, for example, appear to occur after the fusible stimuli have been independently coded by more central processors (see Cutting, 1972, 1973a), suggesting that compatible linguistically coded information, and not compatible raw inputs, interacts to form a higher-level fusion response. In the binaural condition the electrical mixing of the fusible stimuli inhibits this kind of fusion for the synthetic items used in the present study. Their linguistic aspects may have been masked at a lower, more peripheral level so that they reached the higher-level processors in a much degraded form unsuited for fusion.

REFERENCES

Carroll, J. B., P. Davies, and B. Richman. (1971) Word Frequency Book. (New York: Houghton and Mifflin).

Cutting, J. E. (1972) A preliminary report on six fusions in auditory research. Haskins Laboratories Status Report on Speech Research SR-31/32, 93-107.

Cutting, J. E. (1973a) Levels of processing in phonological fusion. Ph.D. thesis, Yale University (Psychology). [Also Haskins Laboratories Status Report on Speech Research SR-34 (this issue).]

Cutting, J. E. (1973b) Phonological fusion in synthetic and natural speech. Haskins Laboratories Status Report on Speech Research SR-33, 19-28.

Cutting, J. E. (1973c) Speech misperception: Inferences about a cue for cluster perception from a phonological fusion task. Haskins Laboratories Status Report on Speech Research SR-33, 57-66.

Cutting, J. E. and R. S. Day. (1972) Dichotic fusion along an acoustic continuum. J. Acoust. Soc. Amer. 52, 175(A). (Also Haskins Laboratories Status Report on Speech Research SR-28, 103-114.)

Day, R. S. (1968) Fusion in dichotic listening. Unpublished Ph.D. thesis, Stanford University (Psychology).

Day, R. S. (1970) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Haskins Laboratories Status Report on Speech Research SR-21/22, 71-87.

Day, R. S. (in preparation) Release from language-bound perception.

Licklider, J. C. R. (1951) Basic correlates of the auditory stimulus. In Handbook of Experimental Psychology, ed. by S. S. Stevens. (New York: John Wiley) 985-1039.

Phonological Fusion of Stimuli Produced by Different Vocal Tracts

James E. Cutting*
Haskins Laboratories, New Haven, Conn.

Phonological fusion is independent of many nonlinguistic constraints that govern other types of auditory fusion (Cutting, 1972). For example, a fusible pair consisting of PAY presented to one ear and LAY presented to the other ear often yields PLAY, whether the two stimuli begin at the same time, share the same fundamental frequency contour, or have the same peak intensity (Cutting, 1973). The present study explores another nonlinguistic dimension--the apparent vocal tract size from which the stimuli were uttered.

Stimuli. Two fusion sets were synthesized on the Haskins Laboratories' parallel resonance synthesizer: the PAY set (PAY, RAY, LAY) and the KICK set (KICK, RICK, LICK). All stimuli were highly identifiable, and both sets were used by Cutting (1973) in observing the effects of other nonlinguistic dimensions on phonological fusion. In the present study stimuli within a set were identical in pitch, intensity, and duration, and differed only in the formant structure of the first 150 msec. The PAY and KICK sets were 350 and 325 msec in duration, respectively. Each item was synthesized in two versions: one as if uttered by a normal adult male with a large vocal tract, and the other as if uttered by a male midget. The small vocal tract stimuli were identical to the large vocal tract stimuli except that the formants were 20 percent higher in frequency, as shown in Figure 1. This change in formant frequency created stimuli that would be uttered by a vocal tract diminished in all dimensions by a factor of 1/6.

Tapes. The stimuli were digitized and stored on disc file for the preparation of dichotic tapes. Dichotic pairs consisted of members of the same stimulus set: for example, PAY/RAY and PAY/LAY. Members of a dichotic pair were uttered by the same vocal tract or by different vocal tracts. For example, "same" pairs were PAY-large/LAY-large (for "large" vocal tract) and PAY-small/LAY-small. "Different" pairs were PAY-large/LAY-small and PAY-small/LAY-large. Two tapes with different random orders included both types of pairs and consisted of 96 dichotic items: (2 sets of stimuli) x (2 stop/liquid pairs per set) x (4 combinations of vocal tract selections) x (3 lead times) x (2 channel arrangements per pair). The lead times selected were the simultaneous onset and 50 msec stimulus onset asynchronies. Two possible 50 msec leads occurred in equal probability with the 0 onset case; for example, LAY began before PAY and PAY began before LAY. The leads were used to make the present study comparable to those in Cutting (1973). Channel arrangements for each pair were counterbalanced.

---

*Also Yale University, New Haven, Conn.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

Figure 1

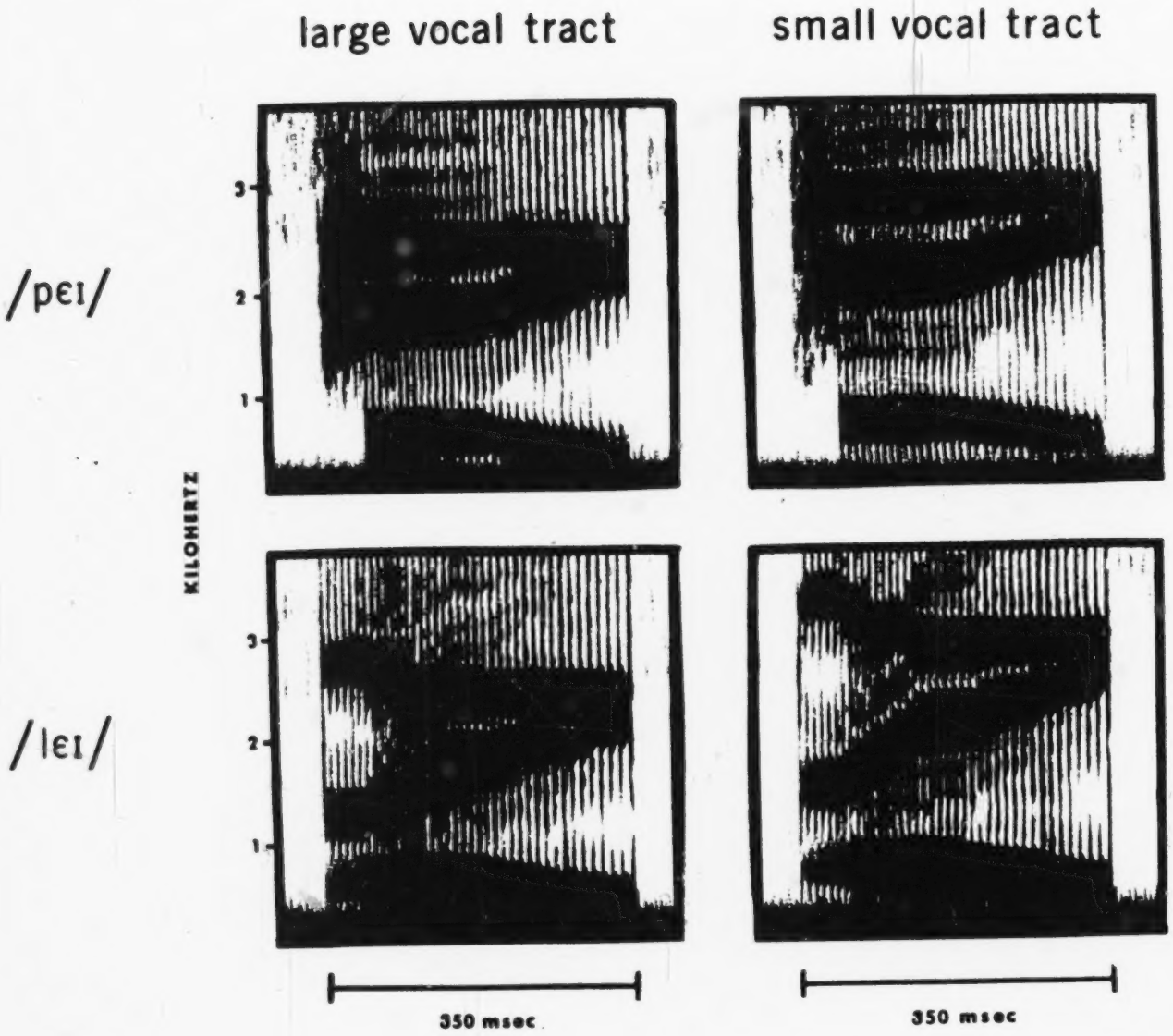


Figure 1: PAY and LAY synthesized as if uttered by vocal tracts of different sizes.

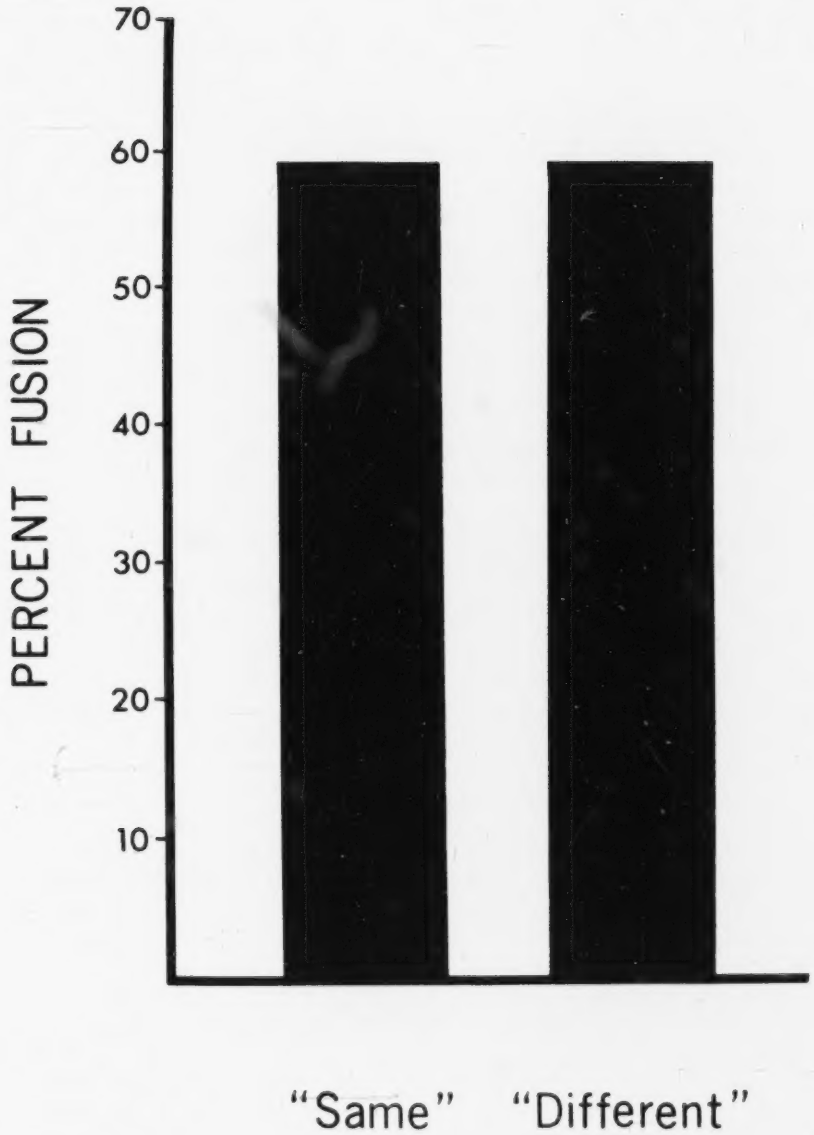


Figure 2: Results of the fusion task when the apparent vocal tract size of the stimuli was varied.

Subjects, apparatus, and procedure. Twelve Yale University undergraduates participated as subjects in this phonological fusion task. Each subject was a right-handed native American English speaker with no history of hearing difficulty. They listened in groups of four to dichotic tapes played on an Ampex AG500 dual track tape recorder sent through an attenuator and listening station to Grason-Stadler earphones (Model TDH39-300Z). Subjects were instructed to write down what they heard: one word or two words, real words or nonsense.

Results. Fusion rate for "same" and "different" pairs was identical--59 percent each, as shown in Figure 2. Furthermore, fusion rate was within a few percentage points for all four combinations of vocal tract selections.

Other results show a pattern similar to that discussed in Cutting (1973). a) Fusion rates for stop + /r/ and stop + /l/ stimuli were nearly identical. b) Many perceptual substitutions occurred among the liquid phonemes: for example, when PAY/RAY fused, PLAY responses were more frequent than PRAY responses. On the other hand, when PAY/LAY fused, PLAY responses occurred nearly all the time. c) Fusions were more frequent when the stop stimulus (e.g., PAY) led the liquid stimulus (e.g., LAY) than in the other two lead-time configurations. d) Individual differences in fusion rate were consistent with previous findings (Day, 1970; Cutting, 1973): some subjects fused at a high rate, others fused at a relatively low rate, and few subjects fused at intermediate rates.

Overview. For phonological fusion to occur, fusible stimuli need not be uttered by the same vocal tract. This result is a further indication that phonological fusion is a higher-level process, not dependent on nonlinguistic compatibility of the stimuli (see Cutting, 1973). Differences in vocal tract size shift the entire formant structure of the to-be-fused stimuli, yet this nonlinguistic variation has little, if any, effect on fusion rate. Fusion appears to occur after the stimuli have been linguistically coded, and this coding process appears to separate linguistic and nonlinguistic information.

REFERENCES

Cutting, J. E. (1972) A preliminary report on six fusions in auditory research. Haskins Laboratories Status Report on Speech Research SR-31/32, 93-107.

Cutting, J. E. (1973) Levels of processing in phonological fusion. Ph.D. thesis, Yale University (Psychology). [Also in Haskins Laboratories Status Report on Speech Research SR-34 (this issue).]

Day, R. S. (1970) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Haskins Laboratories Status Report on Speech Research SR-21/22, 71-87.

Phonetic Prerequisites for First-Language Acquisition*

Ignatius G. Mattingly[+]
Haskins Laboratories, New Haven, Conn.

In a well-known passage in his Aspects of the Theory of Syntax (1965:30), Chomsky lists a number of prerequisites for the infant speaker-hearer's acquisition of competence in his native language. For each of these prerequisites, Chomsky argues, there is an analogous requirement for linguistic investigation. The first prerequisite is "a technique for representing input signals;" that is, the infant, if he is to master his native language, must have at his command an operational universal phonetics. Chomsky's other prerequisites have to do with the infant's capacity to form, test, and select hypotheses about the grammar of his language.

Psycholinguists interested in first-language acquisition have traditionally paid rather less attention to Chomsky's first prerequisite than to the others. The research that has been done on the phonetic capacity of infants and young children has dealt more with the production than with the perception of speech, and has not been very much concerned with what may be called the Representation Problem: how a child "represents input signals." This bias is quite understandable: experimental procedures can be much less sophisticated for study of production than for study of perception. But more recently, a number of experimental studies of infant speech perception have been carried out, using changes in heart rate, sucking rate, or evoked potential to study the infant's ability to recognize and discriminate among speech sounds (see Eimas, in press, for a review).

The Representation Problem is actually only one of the problems that a satisfactory account of the infant's phonetic capacity must solve. It must also solve two other, logically prior problems, which may be called the Speech Detection Problem and the Vocal Tract Problem.

The Speech Detection Problem may at first seem trivial. If the infant is to gather linguistic data from his environment, he must have a way of distinguishing speech sounds from nonspeech sounds. If we choose, of course, we can regard this problem as simply a special case of the more general problem of

---

*Based on talks given at the International Symposium on First-Language Acquisition, Florence, Italy, September 1972, and at a meeting of the Society for Research in Child Development, Philadelphia, Pa., March 1973.

[+]Also University of Connecticut, Storrs.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

pattern recognition by humans, but then, we can also dispose of the whole question of language acquisition in that way. Nor will it do to say that the infant defines speech as those reassuring sounds coming from his mother's mouth (and if he did, we should still have to explain how he arrives at this definition), for many nonspeech sounds come from his mother's mouth, and there are in his environment many other sources of speech, by no means limited to visible mouths. It seems necessary to suppose that the infant has available some procedure that can sort out speech from nonspeech, and we do not know what this procedure is. (For what it is worth, the parallel engineering problem is likewise unsolved: there exists no reliable device or algorithm for automatically sorting speech signals from all nonspeech signals.) In fact, a similar procedure must be attributed to the listener who is already linguistically competent, but the problem appears to be less serious because the competent listener can conceivably apply linguistic and semantic criteria to decide whether he is listening to speech, while the incompetent infant cannot. But we do not know that the mature listener does actually rely primarily on linguistic and semantic criteria, and the fact that he can readily detect speech even when it is unintelligible, or in an unknown language, suggests that he must have phonetic criteria. Almost nothing is known about the Speech Detection Problem, though we shall mention later an observation that sheds a little light on speech detection by both adults and children.

The second problem is the Vocal Tract Problem. How can the infant manage to extract useful linguistic data from the outputs of vocal tracts that, as Lieberman, Crelin, and Klatt (1972) have demonstrated, differ from one another and from his own immature vocal tract in shape, size, and acoustical properties? If we take the position that the linguistic information in speech is represented by acoustic correlates that are, at least at some level of abstraction, invariant, the problem is not so serious: all we have to do is to explain how the infant arrives at this level of acoustic invariance, and then, when he wishes to produce the same sounds, how he works out what the acoustical realization of this invariance would be in his own speech, even though he may never have heard another infant speak. But serious objections have been raised against a view of speech perception that depends on acoustic invariants. We need not go so far as to say that there are no acoustic invariants, but there are not enough of them, and even these available "invariants" sometimes vary. The speech signal simply cannot be analyzed and segmented into units corresponding in any clearcut way with the discrete phonetic elements that the speaker believes he is producing and that the listener believes he is perceiving. Many of us have come to believe that speaker and hearer deal not in acoustic invariants but in articulatory events that are encoded in the acoustic signal by speech cues (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). Thus the fact that /b, d, g/ are respectively labial, alveolar, and velar stop consonants is cued by acoustically observable shifts in the resonances of the vocal tract. These shifts reflect both the stop closure and where in the vocal tract this closure is made. The listener, having tacit knowledge of the properties of the vocal tract--a knowledge of the speech code, in other words--is able to recover the succession of articulatory events. It follows from this view, often called "the motor theory of speech perception" (Liberman, Cooper, Harris, and MacNeilage, 1963), that in order for the infant to gather the data he needs, he, like other speaker-hearers, must know something about the acoustic coding of articulatory events, and he must be able to calibrate his perceptions for a particular vocal tract.

But how is he to learn about vocal tracts? At one time it seemed reasonable to suppose that he learned by studying his own vocal tract and applying some simple mathematical transformation (Fant, 1953). But there are difficulties for this view. Medical cases have been cited of people with congenital gross damage to the vocal tract who do not necessarily have problems in the perception of speech (Lenneberg, 1967). The vocal tracts differ along not just one but a number of partially independent dimensions, e.g., the distance from larynx to palate or the distence from pharyngeal wall to lips. No very simple transformation will suffice. Lieberman et al. (1972) have shown that the vocal tract of an infant is a particularly poor guide to that of an adult: the larynx is too high, the pharynx is too small, and the jaw is too big in proportion. It seems more likely that his own vocal tract is itself a problem for the child. Even if he already has a general knowledge of vocal tracts, he must determine the individual characteristics of his own if he is to produce speech, and perhaps one of the things he is doing during the babbling stage is mapping his vocal tract. Since he can apparently use data from adult vocal tracts to guide production from his own vocal tract, which is not only very different from an adult's but is also changing its configuration rapidly during the period of first-language acquisition, we are forced to say that he must understand not only the physiology of the vocal tract but also something about its ontogeny.

As for our third problem, the Representation Problem, there is both direct and indirect evidence that the capacity to perceive phonetic categories is innate. Abramson and Lisker (1970) have made an extensive cross-language investigation of the acoustic cue of voice onset time (VOT). VOT is the difference in time, positive or negative, between the instant of release of oral closure of stop consonants such as /p/ or /b/ and the beginning of laryngeal voicing. This speech cue occurs in a large number of languages. In English the labial stop will be heard as an aspirated [p^h] if the onset of voicing is delayed as much as 40 msec, but [b] for lower values of VOT. Thus +40 approximates the phoneme boundary that separates /p/ and /b/. In other languages there is a second boundary at about -30 msec; still other languages have three stops at the same position of articulation, and use both phoneme boundaries. Moreover, when subjects are asked to discriminate neighboring sounds along the VOT range, the general finding is that they discriminate extremely well with stimuli close to the phoneme boundaries of their language, and rather poorly elsewhere. Their perception is categorical, as is the case with other speech cues (Liberman, Harris, Hoffman, and Griffith, 1957). The significant point, however, is that there seem to be only two possible VOT phoneme boundaries, regardless of language. This limitation would appear to be a linguistic universal, something that is part of any infant speaker-hearer's innate capacity to acquire language. What he has to learn is whether either or both boundaries are actually used in his native language.

Eimas and his colleagues (Eimas, Siqueland, Jusczyk, and Vigorito, 1971; Eimas, in press) have made a more direct investigation of this question. They tested the ability of four-week-old infants to discriminate VOT differences. Exploiting the fact that their subjects sucked more frequently if presented with a perceptually novel stimulus, they found that the infants could discriminate more readily when the original and the novel stimuli were on opposite sides of the VOT boundary than when both stimuli, though differing by the same amount of VOT, were on the same side of the phoneme boundary. Thus it appears that infants can detect at least one acoustic cue soon after birth. Eimas (in press)

and his colleagues have also studied perception of place of articulation, and have obtained similar results.

Finally, let us return to the Speech Detection Problem. Offhand, one might suppose that there existed somewhere in the auditory system a device for deciding whether what was being heard was nonspeech or speech. To be perceived as speech, the signal would have to satisfy certain criteria of naturalness. If the input signal were judged to be nonspeech, the speech processor system would not be evoked and the information would be sent elsewhere. But if the signal were judged to be speech, some, if not all, of the information in the signal would be sent to the speech processor for further analysis.

This would seem to be a reasonable arrangement, provided we could conceive of a speech detector that was significantly less complex than a speech processor. But there is another possibility that occurred to us because of an unexpected outcome of a recent experiment (Mattingly, Liberman, Syrdal, and Halwes, 1971; see also Eimas, Cooper, and Corbit, in press). Subjects were asked to discriminate a series of very simple synthetic stimuli which were supposed to be speech. An examination of the data, however, revealed fairly clearly that the subjects were hearing the stimuli sometimes as speech and sometimes as nonspeech. Our conclusion was that we had oversimplified our stimuli: they did not contain enough speech cues to sound like speech consistently. This occurrence suggests a different way of looking at the speech detection problem: no speech detector as such is required; rather, speech will be detected provided enough cues are present to arouse the speech processor, even though the stimuli are not very natural sounding. That the brain may work in this way is very fortunate for the experimenter. Like an ethologist, he can present very simplified and hence perhaps very unnatural stimuli to his subjects and yet obtain valid results (Mattingly, 1972). It also suggests that the infant's "technique for representing input signals" is a very robust process that is not easily upset by confusing, inconsistent, or fragmentary input.

REFERENCES

Abramson, A. S. and L. Lisker. (1970) Discriminability along the voicing continuum: Cross-language tests. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. (Prague: Academia).

Chomsky, N. (1965) Aspects of the Theory of Syntax. (Cambridge, Mass.: MIT Press).

Eimas, P. D. (in press) Speech perception in early infancy. In Infant Perception, ed. by L. B. Cohen and P. Salapatek. (New York: Academic Press).

Eimas, P. D., W. E. Cooper, and J. D. Corbit. (in press) Some properties of linguistic feature detectors. Percep. Psychophys.

Eimas, P. D., E. R. Siqueland, P. Jusczyk, and J. Vigorito. (1971) Speech perception in infants. Science 171, 303-306.

Fant, C. G. M. (1953) Comment on paper by G. Peterson. In Communication Theory, ed. by W. Jackson. (New York: Academic Press).

Lenneberg, E. (1967) Biological Foundations of Language. (New York: Wiley).

Liberman, A. M., F. S. Cooper, K. S. Harris, and P. F. MacNeilage. (1963) Proceedings of the Speech Communication Seminar, Stockholm, 1962.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith. (1957) The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 54, 358-368.

Lieberman, P., E. S. Crelin, and D. Klatt. (1972) Phonetic ability and related anatomy of the newborn and adult human, Neanderthal man, and the chimpanzee. Amer. Anthropol. 74, 287-307.

Mattingly, I. G. (1972) Speech cues and sign stimuli. Amer. Scient. 60, 327-337.

Mattingly, I. G., A. M. Liberman, A. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. 2, 131-157.

A Note on the Relation between Action and Perception*

M. T. Turvey+
Haskins Laboratories, New Haven, Conn.

I would like to explore the thesis that, in principle, the fundamental problems to be solved and the concepts that will provide their solution are the same for both the Theory of Action and the Theory of Perception. Consider the following transcription task. A person sees a written capital A and is required to respond by writing the letter that she has seen. There can be, of course, different tokens of the letter seen, and there can be a variety of manners in which our person is requested to make her written response. We may partition a perceptual-motor occurrence of this kind into three phrases. The first consists of a set of functions that map states of the optic space ($\underline{o}$) into states of the perceptual space ($\underline{p}$); the second consists of a set of functions that map states of the perceptual space into states of the act space ($\underline{a}$); and the third is a set of functions that map states of the act space into states of the motor space ($\underline{m}$). We can represent the three phases as: $F_1 = \{f|f:\underline{o}\rightarrow\underline{p}\}$, $F_2 = \{f|f:\underline{p}\rightarrow\underline{a}\}$, and $F_3 = \{f|f:\underline{a}\rightarrow\underline{m}\}$.

The thesis to be explored suggests that we should look for similarities between the functions on the perceptual end and those on the action end of our transcription task, i.e., between F_1 and F_3. In addition, it suggests that we should ask in what way(s) the perceptual and act spaces may be similar. The present paper is a response to these suggestions.

The concept of action

I am going to assume that our collective intuitions about the concept of perception will probably suffice for the present purposes. However, I will not make the same assumption for the concept of action, primarily because there has been far less hue and cry about this concept in theoretical psychology. Consequently, its character is less well articulated--witness the tendency to equate action with response in comparison to the tendency to equate perception with stimulus. We do the latter rarely, and the former frequently. Action, like perception, is an abstract relation between the organism and the environment. Just

---

*Paper presented at the Allerton Conference of the North American Society for the Psychology of Sport and Physical Activity, Monticello, Ill., 13-17 May 1973.

+Also University of Connecticut, Storrs.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

as we are unable to point to perception, we are unable to point to action, although, of course, we might be able to detail the parameters of stimulation, and to describe fully the muscles and joints involved in an exhibited movement.

A number of philosophers (see Care and Landesman, 1968) have sought to lay bare the concept of action. Here I will report only briefly on their endeavors merely to portray two rather important characteristics of the concept. The first is that nerve impulses and muscle contractions--though necessary conditions for action--are more accurately considered accompaniments of action than characteristics of action. This point can be defended on at least two grounds: first on the notion of intentionality, and second (to be explored at some length below) on the issue of movement equivalence, or constancy. Clearly, it is perfectly reasonable to say that one intends to kick a football, but it is not reasonable to say that one intends to contract and to relax one's biceps femoris and one's rectus femoris, respectively, to this or that degree. Generally one cannot choose or intend to transmit a nerve impulse or to contract a certain set of extrafusal and intrafusal muscle fibers. Intentionality is a defining characteristic of acts, but not of muscle contractions.

Another defense of the notion that the concept of action cannot be reduced to bodily movement is that any particular constellation of muscular contractions and joint motions brought into play when one performs an act (say, reaching for and lifting up a cup) cannot be said to be identical to the act. A radically different configuration of muscles and joints could just as easily have been used to achieve the same result. An act is Gestalten, that is to say, in a variant of the hackneyed phrase of Gestalt Psychology: an act is more than the sum of its constituent movements.

The second important aspect of the concept of action is that consequences, i.e., changes wrought in the environment by a configuration of movements, are integral to the concept of action in that no reliable distinction can be drawn between the concepts of action and consequence. Consider the following: George kicks the football (of the round kind), and scores the goal that wins the championship. Now we could say that George kicked the football and that a consequence of this action was that a goal was scored. Or, we could say, just as appropriately, that George scored a goal with championship-winning consequences. "Scores the goal," therefore, can be viewed either as consequence or as action. We might suppose that there are criteria available to determine what occurrences should receive an action label, and what occurrences should receive a consequence label. Unfortunately, the criteria that have been advanced have not been greeted with universal approval.

The problem of constancy in perceiving and acting

Let us now return to the phases in transcription referred to above, in particular F_1 and F_3. Consider that a visually presented capital A can occur in various sizes and orientations and in a staggering variety of individual scripts. Yet in the face of all this change, the identification of the letter remains constant; we see through the variations to the canonical form.

This phenomenon of constancy is not limited to the domain of perception, but is equally characteristic of action. Thus, the letter A may be written without moving any muscles or joints other than those having to do with the fingers. Or,

it may be written through large movements of the whole arm with the muscles of the fingers serving only to grasp the writing instrument. Or, more radically, one can write the character without involving the muscles and joints of either arms or fingers, by clenching the writing instrument between one's teeth or toes. It is evident that a required result can be attained by an indefinitely large class of movement patterns.

On examination of the phenomenon of constancy we might raise the query: How can these indefinitely large classes of possible shapes, and of possible movement patterns, be stored in memory? The answer is that they are not. Clearly, I do not have on record in memory all possible visual versions of A, since I have never experienced most of them. And similarly, I do not have memorized all possible temporal sequences of all possible configurations of muscle motions that write A; indeed, I have yet to perform them and by all accounts I never will. The essential question about our transcription task, therefore, can be stated more fundamentally: How can I recognize and produce the indefinitely various instantiations of A without previous experience of them?

In response to this question let us turn our attention to linguistic theory. The point of departure for transformational grammar is that our competency in language is such that we can produce and understand a virtually infinite number of sentences. As Weimer (1973) has pointed out, there are echoes of Plato's paradoxes in Chomsky's (1965) claim that our competence in language vastly outstrips our experience with it. Chomsky's claim is motivated by the observation that experience with a limited sample of the set of linguistic utterances yields an understanding of any sentence that meets the grammatical form of the language. To explain this competency is, for Chomsky (1966), a central problem in the Theory of Language. But given the points advanced above, the constancy function in action and perception is likewise indicative of a competency that exceeds prior learning. The child, we may note, learns to write A under conditions which restrict her to a small subset of the very large set of A-writing movements. But she is able subsequently to write A with practically any movement pattern she chooses, i.e., she can write A in novel ways. Similarly, limited visual experience with some A's is sufficient to allow the child to identify virtually any A. Thus, acting and perceiving are creative in the sense that language is creative, and I would submit, therefore, that the explanation of this creativity is central to the theories of action and of perception, and at the very heart of our understanding of perceptual-motor skill.

The search for a workable account of the creativity manifest in language has led transformational grammarians to what has been aptly described as "the explanatory primacy of abstract entities" (Weimer, 1973). The idea is that the speaker-listener has at his disposal an abstract system of rules or principles referred to as the deep structure that allows him to generate and to understand an indefinitely large set of sentences referred to as the surface structure. This distinction, drawn in linguistic theory, between deep and surface structure applies to our present concerns in two important respects. The first is the transformational grammarian's view that deep structure is far removed from surface structure; it is argued that although the deep structure determines the surface structure it is not manifested in the surface structure. The importance of this view is that it concurs with Bernstein's (1967) general conclusion in his classic analysis of the coordination and regulation of movement. Referring to the engram or motor-image of an act Bernstein comments: "The higher engram,

which may be called the engram of a given topological class, is already structurally far removed from any resemblance whatever to the joint-muscle schemata...." (p. 49). The essence of Bernstein's (1967) view is that the central substrate for a pattern of movements is a representation of the environment. Following Evarts' (1967) work, Pribram (1971) has argued that the cortical representation can be thought of as "a 'mirror image' of the field of external forces" (p. 246). Thus, the underlying structure is best described as an Image-of-Achievement since it encodes environmental contingencies. An interesting upshot of this view is that action and consequence, which prove to be inseparable conceptually, are also inseparable neurophysiologically.

The second characteristic of the surface-deep structure distinction I wish to touch upon is that the child must come to determine the nature of the underlying deep structure from a limited experience with surface structures. It is assumed by Chomsky and his colleagues that the child essentially "looks through" the utterances she hears to the abstract form behind those utterances. The child is said, therefore, to construct a theory of the regularities of her linguistic experience. Similarly, the child seeing capital A's must determine an abstract representation that will extend over an indefinitely large set of instantiations of that character. And, by the same token, our hypothetical child learning to write the letter A must determine from her limited experience with the set of A-writing movements a theory of how to write A. The abilities to recognize indefinitely various A's, and to write A in indefinitely various ways are based on representations that are abstract and generative, like the grammar Chomsky has in mind for language. We should not be surprised by this conclusion: there is no reason why the nervous system should not solve similar problems in similar ways.

The mathematical group as an example of an abstract structure

Clearly, the form of the representation that allows for the writing of A in novel ways is not motor. That is, it cannot be said to consist of programs of muscle innervation. In the same way, the abstract representation that affords the identification of novel A's cannot be sensory, i.e., it cannot be described as any circumscribed set of sensory properties. We should note that the constancy function in the identification and in the writing of A reveals an indifference of both modes to metrical variation, and suggests rather strongly a dependency of both modes on topological properties (Bernstein, 1967). Thus, common to all capital A's viewed and written is that they are members of a single topological class, while the differences between capital A's, both viewed and written, would be determined by topological differences of a higher order (Bernstein, 1967).

On the foregoing considerations we should argue that the action concept of A and the perception concept of A cannot be represented, respectively, as a particular aggregate of motor elements and as a particular aggregate of sensory elements. Instead, they are more accurately viewed as injunctions or rules specifying how a set of elements should relate, whatever those elements might be.

In attempting to account for constancy in visual perception several students of the problem have appealed to the mathematical concept of group (e.g., Cassirer, 1944; Pitts and McCulloch, 1947). Essentially, a group is any set or collection of elements (and they need not be specified) which can be combined according to a

law such that any combination of them produces an element belonging to the set itself. The set, therefore, is said to be self-contained or closed.

More formally, a group may be defined as a set G together with a composition rule which generates for each pair a and b of elements of G a third element ab of G for which the following conditions hold:

1. The composition rule is associative: For any three elements a, b, c of G: $(ab)c = a(bc)$.
2. There exists an element i in G such that $a \cdot i = i \cdot a = a$. The element i is known as the identity element.
3. To each element in G there corresponds an inverse in G such that: $a \cdot a^{-1} = a^{-1} \cdot a = i$.

If we now define a generic concept as a group of transformations (Cassirer, 1944), then it can be said that there is a group G, which defines the action concept of A and another group g, which defines the perception concept of A. However, an interesting property of groups is that two groups can be isomorphic, that is, they can represent the same abstract group, if the manner of internal interlocking of elements is the same in both cases, even though the elements of the two groups may differ radically from one another in other respects. Thus, although the perception concept of A and the action concept of A would appear to differ because of the different elements with which they work (sensory properties on the one hand and muscle contractions on the other) they may, in fact, be identical. The idea is that the two groups, G and g, which define the two concepts, have the same internal structure. This speculative conclusion can be stated more usefully as follows: the abstract structure that affords the identification of optical instantiations of A also affords the production of motor instantiations of A.

Conclusion

I have considered certain characteristics of a "simple" transcription task in order to argue that the problems that beset the perception theorist and those that beset the action theorist are very similar in nature, and thus similar in the principles needed for their solution. On a less general level I have speculated that for this transcription task (and I suspect for others) perception and action maybe related through a common abstract structure indigenous to neither. And finally, I have expressed, implicitly, the view that the Theory of Action is as much a part of Cognitive Psychology as are the Theory of Perception and the Theory of Language. If you remain unconvinced of the abstract nature of the knowledge underlying so-called perceptual-motor activity, consider the following description of balancing on a bicycle presented by Michael Polanyi (1964). As the cyclist starts to fall to the right he turns the handlebars to the right, deflecting the bicycle along a curve to the right. The result of this maneuver is a centrifugal force pushing the cyclist to the left and offsetting the gravitational force pulling to the ground to the right. Consequently the cyclist is thrown out of balance to the left and responds by turning the handlebars to the left deflecting the bicycle along a curve to the left, which results in a centrifugal force pushing him to the right, etc., etc. In the course of these maneuvers the cyclist is obeying the following injunction: "adjust the curvature of the bicycle's path in proportion to the ratio of unbalance over the square of the speed." Keeping one's balance on a bicycle is a very cognitive act.

REFERENCES

Bernstein, L. (1967) The Coordination and Regulation of Movements. (London: Pergamon Press).

Care, N. S. and C. Landesman. (1968) Readings in the Theory of Action. (Scarborough, Ont.: Fitzhenry and Whiteside).

Cassirer, E. (1944) The concept of group and the theory of perception. Philosophy and Phenomenological Research 5, 1-36.

Chomsky, N. (1965) Aspects of the Theory of Syntax. (Cambridge, Mass.: MIT Press).

Chomsky, N. (1966) Topics in the Theory of Generative Grammar. (The Hague: Mouton).

Evarts, E. V. (1967) Representation of movements and muscles by pyramidal tract neurons of the precentral motor cortex. In Neurophysiological Basis of Normal and Abnormal Motor Activities, ed. by M. D. Yahr and D. P. Purpura. (New York: Raven Press).

Pitts, W. H. and W. S. McCulloch. (1947) How we know universals: The perception of auditory and visual forms. Bull. Math. Biophys. 9, 127-147.

Polanyi, M. (1964) Personal Knowledge: Towards a Post-Critical Philosophy. (New York: Harper).

Pribram, K. H. (1971) Languages of the Brain. (Englewood Cliffs, N. J.: Prentice Hall).

Weimer, W. B. (1973) Psycholinguistics and Plato's paradoxes of the Meno. Amer. Psychol. 28, 15-33.

Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing*

David B. Pisoni[+] and Jeffrey Tash[+]

Same-different reaction times (RTs) were obtained in a Posner-type matching task to pairs of synthetic speech sounds ranging perceptually from /ba/ through /pa/. Listeners were required to respond "same" if both stimuli in a pair were the same phonetic segments (i.e., /ba/ - /ba/ or /pa/ - /pa/) or "different" if both stimuli were different phonetic segments (i.e., /ba/ - /pa/ or /pa/ - /ba/). RT for "same" matches was faster to pairs of acoustically identical stimuli (A-A) than to pairs of acoustically different stimuli (A-a) belonging to the same phonetic category. RT for "different" responses was slower for a two-step difference across the phonetic boundary than for a four-step or six-step difference. These results provide evidence for distinct auditory and phonetic levels of processing in speech perception. Low-level acoustic information about stop consonants may be available to listeners but this is dependent on the level of processing accessed by the particular information processing task employed.

It is now well established that when listening in the speech mode a subject can identify the phonetic category of a sound but cannot discriminate between acoustically different sounds selected from within the same phonetic category (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967; Liberman, 1970; Pisoni, 1971, 1973). This phenomenon, known as "categorical perception," appears to be unique to certain classes of speech sounds. In the idealized case, two speech sounds can be discriminated only to the extent that they can be identified as different on an absolute basis (Studdert-Kennedy, Liberman, Harris, and Cooper, 1970). This contrasts with other kinds of auditory perception where discrimination is better than absolute identification (Pollack, 1952, 1953).

---

*An earlier report of these findings was presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., 11 April 1973.

[+]Indiana University, Bloomington.

Acknowledgment: This research was supported by a grant from NICHD to Haskins Laboratories and a PHS grant (S05 RR 7031) to Indiana University. We are very grateful to Professor Lloyd Peterson for the use of his computer and to R. M. Shiffrin, M. Studdert-Kennedy, and I. Pollack for their interest and advice on this project.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

Auditory information from the earliest stages of perceptual processing of speech sounds may be lost as a consequence of phonetic categorization. Thus, acoustic information will be unavailable for use in a subsequent discrimination task (Pisoni, 1973). The complexity of speech sounds and the status they have as linguistic segments in language may force listeners to respond to these sounds in an absolute sense, transforming the sounds into more durable phonetic representations. Since the discrimination tasks employed in most speech perception experiments place a heavy load on short-term memory, it is reasonable to suppose that a phonetic representation would be stored in short-term memory in preference to the auditory transform of the complex acoustic signal. Accordingly, the observed "categorical" discrimination performance may not actually be based on the specific acoustic properties of the stimuli, but rather, on a higher, more abstract phonetic level of analysis. It is possible that information from the earliest stages of processing might be available to a listener, at least for a short period of time. However, the extent to which this relatively unencoded, low-level information can be accessed will depend on a variety of different factors including the stage or stages of perceptual analysis examined by a particular information processing task.

The present study is concerned with how listeners go from one level of perceptual analysis to another in speech perception and with what type of information may remain of previous levels of analysis. Specifically, we were concerned with determining whether listeners could respond to acoustic differences between categorically perceived speech sounds or whether they can only process these sounds on an abstract phonetic basis. The procedure used to investigate this problem was the reaction time (RT) matching paradigm developed by Posner and his associates (Posner and Mitchell, 1967; Posner, 1969; Posner, Boies, Eichelman, and Taylor, 1969). This procedure provides an opportunity to examine the levels of analysis at which comparisons are made by measuring the processing time required for different types of comparisons.

Thus, when a listener is asked to determine whether two speech sounds are the "same" or "different," the time to arrive at a decision may reflect the level of perceptual processing and in turn the type of information required for a comparison. Some speech sounds may be compared directly, based on their acoustical properties, while other stimuli may require a process of abstraction where invariant features must first be identified before being compared (Posner and Mitchell, 1967; Posner, 1969). Classifying two <u>acoustically different</u> speech sounds as the "same" may be considered to involve matching abstracted phonetic features at a higher level of perceptual analysis than classifying two <u>acoustically identical</u> stimuli as the "same." The latter comparison could be based on an earlier stage of analysis involving only the low-level acoustic properties of the stimuli.

Figure 1 shows a flowchart of a model of the stages of analysis involved in this type of classification task. This model is adapted from Posner and Mitchell's (1967) work on letter classification.

On every trial a listener is presented with a pair of stimuli and is required to determine whether the members of the pair are the "same" or "different." Three types of stimulus pairs are shown at the top of the figure, A-A, A-a, and A-B. The A-A pairs represent acoustically identical pairs of stimuli. The A-a pairs represent acoustically different stimuli selected from within a

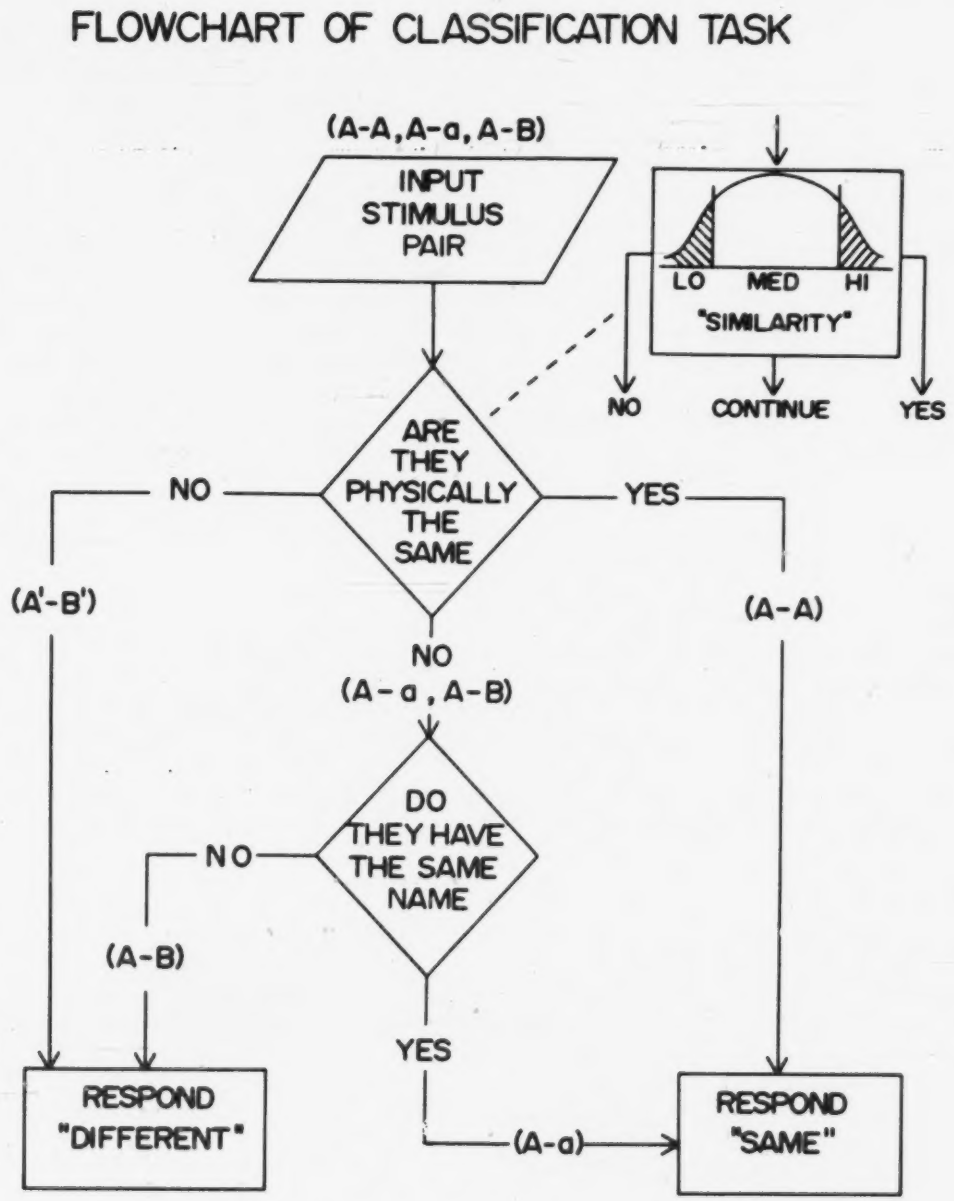


Figure 1: Model of the stages of analysis involved in the "same" - "different" classification task.

particular phonetic category. [In Posner's (1969) terminology, these would be pairs of physically different stimuli with the same "name" code.] Finally, the A-B pairs represent stimuli selected from different phonetic categories. These are acoustically different and have different names.

Depending on the particular type of stimulus pair, various predictions can be made about the relative amount of time required for "same" matches and "different" matches. For example, if low-level acoustic information can be accessed for a comparison, reaction time should be faster for a "same" response when the input pairs are acoustically identical (e.g., A-A) than when they are acoustically different but phonetically the same (e.g., A-a). This should be true if the acoustically identical pairs (e.g., A-A) could be matched as "same" at an earlier stage of analysis than the acoustically different pairs (e.g., A-a). The acoustically different pairs would require an additional stage of analysis for a "same" response. However, if only an abstract phonetic representation is used in the comparison, reaction times for a "same" match to these two types of pairs should be identical. Under this assumption, a similar set of predictions can also be made for the "different" matches. If distinct auditory and phonetic levels of processing exist, pairs of stimuli with large physical differences should be matched as "different" faster than pairs of stimuli with smaller physical differences. If only an abstract phonetic representation is available, reaction time for "different" matches should be equivalent, regardless of the magnitude of the physical differences between pairs of stimuli.

METHOD

Subjects

The listeners were nine paid volunteers, all of whom were either graduate students or staff members associated with the Mathematical Psychology Program at Indiana University. The Ss were right-handed native speakers of English and reported no history of a hearing disorder or speech impediment. Ss were paid for their services at the rate of $1.50 per hour. All Ss had had some previous experience with synthetic speech stimuli, although they were naive to the exact purposes of the present experiment.

Stimuli

A set of bilabial stop consonant-vowel (CV) stimuli were synthesized on the parallel resonance synthesizer at Haskins Laboratories. The basic set of stimuli consisted of seven three-formant syllables 300 msec in duration. The stimuli varied in 10-msec steps along the voice onset time (VOT) continuum from 0 through +60 msec, which distinguishes /ba/ and /pa/. VOT has been defined as the interval between the release of the articulators and the onset of laryngeal pulsing or voicing (Lisker and Abramson, 1964). Synthesizer control parameter values for these stimuli were similar to those employed by Lisker and Abramson (1970) in their cross-language experiments. The final 250 msec of the CV syllable was a steady-state vowel appropriate for an English /a/. The frequencies of the first three formants were fixed at 769, 1,232, and 2,525 Hz respectively. During the initial 50-msec transitional period, the first three formants moved upward toward the steady-state frequencies of the vowel. For successive stimuli in the set, the delay in the rise of F1 to full amplitude (i.e., the degree of F1 "cutback") and in the switch of the excitation source

from hiss (aperiodic) to buzz (periodic) was increased by 10 msec. Simultaneous changes in amplitude in the lower frequency region and type of excitation source have been shown to characterize the voicing and aspiration differences between /b/ and /p/ in English (Lisker and Abramson, 1967).

Experimental Materials

All stimuli were digitized and their wave forms stored on the Pulse Code Modulation System at Haskins Laboratories (Cooper and Mattingly, 1969). Two types of audio tapes were prepared under computer control: an identification test and a matching test. A 1,000 Hz tone of 100 msec duration was recorded 500 msec before the onset of each trial. This tone served as a warning signal for the S and was also used to trigger a computer interrupt which initiated timing response latency.

Two different 140-trial identification tests were prepared. Each test contained 20 different randomizations of the seven stimuli. Stimuli were recorded singly with a 3-sec interval between presentations. Each stimulus occurred equally often within each half of the tape.

Four different "same" - "different" matching tests were constructed. Each test tape contained 48 pairs of stimuli. Half of all the trials consisted of within-category pairs requiring a "same" response while the other half consisted of across-category pairs requiring a "different" response. Figure 2 shows the arrangement of the stimulus conditions employed in the present experiment.

Within-category pairs were either physically identical (A-A) or physically different (A-a). A-A trials consisted of stimuli 1, 3, 5, and 7, each paired with itself (i.e., 1-1, 3-3, 5-5, 7-7). The A-a trials, which were separated by two steps along the continuum or +20 msec VOT, consisted of the stimulus pairs 1-3, 3-1, 5-7, and 7-5.

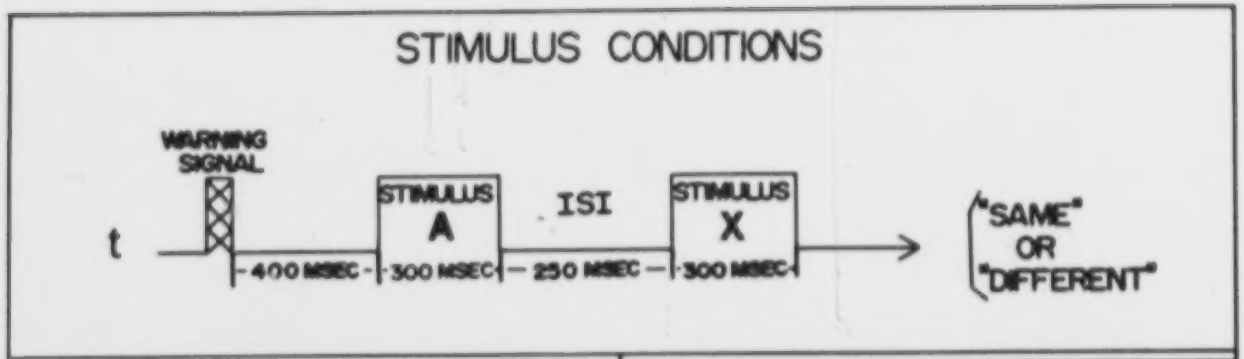
Across-category pairs (A-B), which were always physically different, were separated by two, four, or six steps along the continuum. These comparisons represented differences of +20, +40, or +60 msec VOT respectively.

Each of the eight within-category comparisons appeared three times within a block of 48 trials, whereas each of the six between-category comparisons appeared four times. The interstimulus interval (ISI) between members of a pair was held constant at 250 msec. Successive trials were separated by 4 sec.

Procedure

The experimental tapes were reproduced on an Ampex AG-500 two-track tape recorder and were presented diotically through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the earphones equivalent to 70 db SPL re 0.0002 dynes/cm^2 for a 1,000 Hz calibration tone. Measurements were made on a Hewlett Packard VTVM (Model 400) before the presentation of each experimental tape. Ss were run individually in a small experimental room. All responses and reaction times were recorded automatically under the control of a PDP-8 computer located with the tape recorder in an adjacent room.

Figure 2



WITHIN CATEGORY PAIRS "SAME RESPONSE"		ACROSS CATEGORY PAIRS "DIFFERENT RESPONSE"		
A-A (0-STEP)	A-a (2-STEP)	A-B (2-STEP)	A-B (4-STEP)	A-B (6-STEP)
1-1	1-3	3-5	2-6	1-7
3-3	3-1	5-3	6-2	7-1
5-5	5-7			
7-7	7-5			

Figure 2: Description of the stimulus conditions employed in the matching task. Pairs of stimuli requiring a "same" response are selected from within a phonetic category; pairs of stimuli requiring a "different" response are selected from across phonetic categories.

The instructions for the identification test were similar to those used in previous speech perception experiments. Ss were required to identify each stimulus as either /ba/ or /pa/ and to respond as rapidly as possible. The Ss responded to each stimulus by pressing one of two labeled telegraph keys. For a given S, one hand was always used for a /ba/ response while the other hand was used for a /pa/ response. The two keys were counterbalanced for hands across Ss.

For the matching task, Ss were told that they would hear a pair of stimuli on every trial and their task was to decide whether the two stimuli were the "same" or "different" phonetic segments. The type of instructions employed here is similar to the "name match" instructions employed by Posner and Mitchell (1967). Ss were told that half of all the pairs were the same (e.g., /ba/ - /ba/ or /pa/ - /pa/) and half of the pairs were different (e.g., /ba/ - /pa/ or /pa/ - /ba/). Ss were encouraged to respond as rapidly as possible. As in the identification task, Ss responded to each pair by pressing one of two telegraph keys, labeled "same" and "different." The response keys were also counterbalanced for hands across Ss.

Ss were tested for an hour a day on two consecutive days. Each session began with a 140-item identification test which was followed after a short break by two 48-trial matching tests. Since the first day served as a practice session, only the identification and matching data from the second session will be considered in the remainder of this report.

RESULTS AND DISCUSSION

Identification Task

The average identification function is shown in Figure 3 along with the mean RT for identification. Each point represents the mean of 180 responses over the nine Ss.

The filled squares and open circles show percent /ba/ or /pa/ response respectively to each of the seven stimuli in the continuum. The filled triangles represent the corresponding latency of identification response to each stimulus. Examination of Figure 3 indicates that the identification function is quite consistent. Ss partitioned the stimulus continuum into two discrete phonetic segments. The phonetic boundary or cross-over point in identification is at about +30 msec VOT which is consistent with previous findings (Lisker and Abramson, 1967).

Inspection of the RT function during identification shows that Ss are slowest for stimulus 4 which is at the phonetic boundary and fastest for the other stimuli which are within phonetic categories. These results are also consistent with the findings reported by other investigators who have studied reaction time in the identification of synthetic speech sounds (Studdert-Kennedy, Liberman, and Stevens, 1963). Reaction time is a positive function of uncertainty, increasing at the phonetic boundary where identification is least consistent and decreasing where identification is most consistent. In anticipation of the discrimination tests, it is noted that identification time is slowest for the stimulus region where discrimination is best.

Figure 3

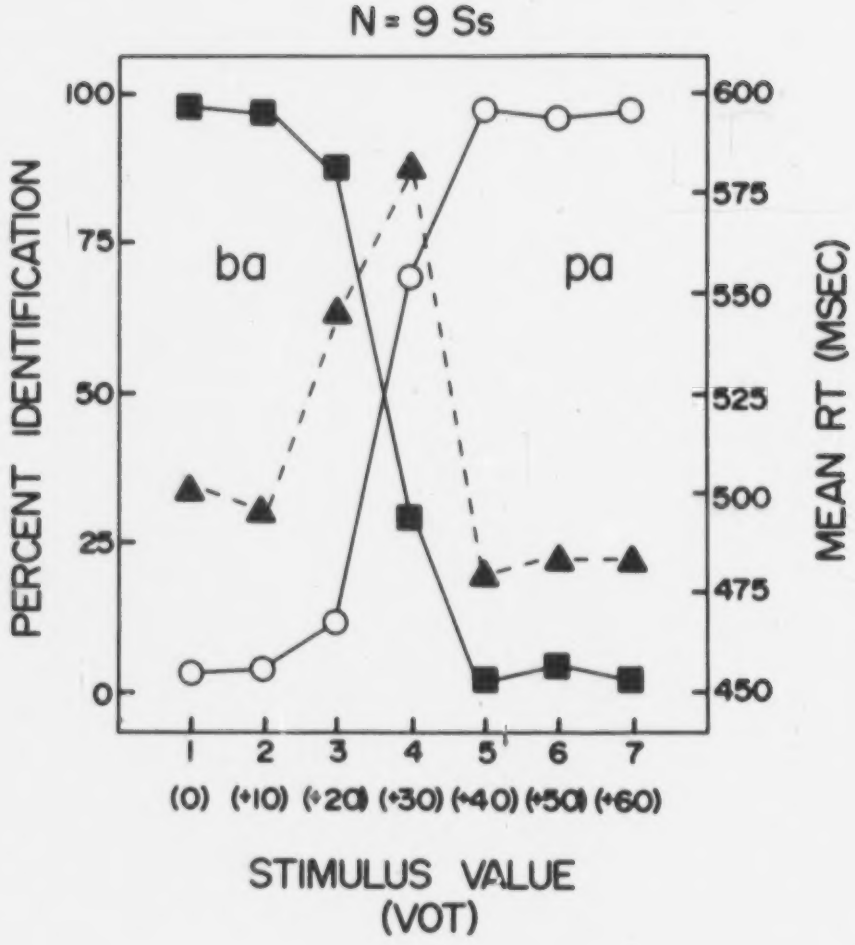


Figure 3: Average identification function for the voice onset time continuum with mean RT dur'ng identification.

Matching Task

The major results of the "same" - "different" classification task are shown in Figure 4.

The mean RT for each of the two types of "same" trials (A-A, A-a) is based on a total 216 judgments, while the mean RT for each of the three types of "different" pairs is based on 144 judgments averaged over nine Ss.

An examination of the "same" responses reveals that subjects are faster for pairs of acoustically identical stimuli (e.g., A-A) than for pairs of acoustically different stimuli (e.g., A-a) which have been selected from within a phonetic category. The 41 msec difference between these two conditions is highly significant ($P<.01$) by a correlated t-test, $t(8) = 3.20$ (one-tailed). This result is consistent with the model described earlier. Ss can access low-level acoustic information even though they have categorized these pairs of stimuli as the "same" phonetic segments. Thus, "same" matches to acoustically identical speech sounds are presumably based on an earlier stage of perceptual analysis than "same" matches to acoustically different speech sounds. In the latter case, the "same" response is based on a comparison of the phonetic features of each stimulus which must have been extracted before a match could have been made. It may be assumed that the abstraction of phonetic features from the acoustic signal requires an additional amount of processing time. This is presumably responsible for the difference in RT between the two within-category conditions.

The present findings, based on "same" responses to within-phonetic category comparisons indicate that even perception of stop consonants may not be entirely categorical, as previously supposed. Rather, the degree of categorical perception will depend upon the extent to which low-level acoustic information can be utilized within the experimental task. Since acoustic information not only decays rapidly over time but also is highly vulnerable to various types of interfering stimuli, the specific discrimination procedure may be crucial in determining the relative roles of acoustic and phonetic information in speech sound discrimination. For example, the ABX procedure may force the listener to rely almost entirely on phonetic information in discrimination because of the arrangement of stimuli in this procedure.

One additional point should be emphasized here concerning the within-phonetic category comparisons. It could be argued that, in the present experiment, Ss were responding to these stimuli as isolated acoustic events rather than as speech sounds. If so, the proportion of "same" responses should have been quite different for A-a pairs and A-A pairs. In fact, P('same'|A-a) and P('same'|A-A) were almost identical, suggesting that Ss were responding to these stimuli as speech sounds.

The mean RTs for "different" responses to the three types of across-category pairs are also shown in Figure 4. An examination of these RTs provides additional support for the argument that Ss can employ relatively low-level acoustic information in the comparison process. RT for "different" matches is slower for a two-step difference than for a four-step or six-step difference across the phonetic boundary. Both differences are significant ($p<.005$) by correlated t-tests, $t(8) = 4.95$; $t(8) = 4.82$, respectively. These findings

Figure 4

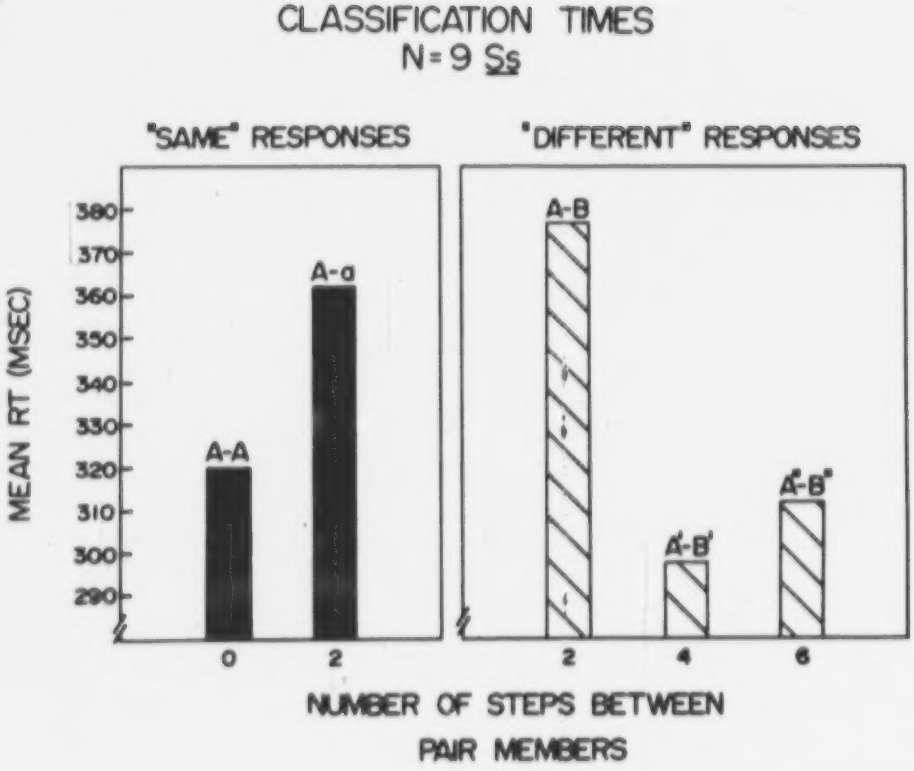


Figure 4: Mean RT for "same" and "different" responses to within- and across-category comparisons. The number of steps between pair members reflects the magnitude of the acoustic difference in voice onset time (VOT).

suggest that "different" matches may not be based solely on an abstract phonetic representation. Rather, a "different" response to pairs of stimuli across category boundaries may also be based on low-level acoustic information at an earlier stage of perceptual analysis. Pairs of stimuli which are separated by large physical differences in VOT, such as 1-7 and 2-6, can be differentiated solely on the basis of their acoustic dissimilarity. Stimulus pairs separated by small differences in VOT, such as 1-3 and 3-5, cannot be differentiated on the basis of their acoustic similarity and an additional stage of analysis is required. Since an initial decision cannot be made reliably on the basis of acoustic information alone, the "different" decision for pair 3-5 must, therefore, be based on a comparison of the phonetic features of the two stimuli.

In summary, the results suggest that low-level acoustical information about a speech stimulus may be available to listeners along with a more abstract phonetic representation, even in the case of stop consonants. Presumably the extent to which low-level information can be accessed will depend not only on the particular level of perceptual analysis examined but also on the type of information processing task employed.

The results of this experiment argue for a diversity of experimental tasks in the study of speech sound perception. On the basis of the distribution of responses alone, we might conclude that only a categorical or phonetic analysis is available for stop consonants. The addition of the RT task reveals another level of analysis. A view of speech sound perception entailing a series of interrelated stages of analysis could serve as the framework for determining quantitatively the ways in which speech perception may involve specialized mechanisms for perceptual analysis. Moreover, such an approach may help to determine the ways in which various speech perception phenomena may conform to more general perceptual processes.

REFERENCES

Cooper, F. S. and I. G. Mattingly. (1969) Computer-controlled PCM system for investigation of dichotic speech perception. Haskins Laboratories Status Report on Speech Research SR-17/18, 17-21.

Liberman, A. M. (1970) Some characteristics of perception in the speech mode. In Perception and Its Disorders, Proceedings of A. R. N. M. D., ed. by D. A. Hamburg. (Baltimore, Md.: Williams and Wilkins) 238-254.

Liberman, A. M., F. S. Cooper, D. S. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Lisker, L. and A. S. Abramson. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384-422.

Lisker, L. and A. S. Abramson. (1970) The voicing dimension: Some experiments in comparative phonetics. In Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. (Prague: Academia) 563-567.

Pisoni, D. B. (1971) On the nature of categorical perception of speech sounds. Haskins Laboratories Status Report on Speech Research SR-27, 101.

Pisoni, D. B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. Psychophys. 13, 253-260.

Pollack, I. (1952) The information in elementary auditory displays. J. Acoust. Soc. Amer. 24, 745-749.

Pollack, I. (1953) The information in elementary auditory displays II. J. Acoust. Soc. Amer. 25, 765-769.

Posner, M. I. (1969) Abstraction and the process of recognition. In The Psychology of Learning and Motivation, ed. by G. H. Bower and J. T. Spence. (New York: Academic Press) 44-100.

Posner, M. I., S. J. Boies, W. H. Eichelman, and R. L. Taylor. (1969) Retention of visual and name codes of single letters. J. Exp. Psychol. Monogr. 79, 1-16.

Posner, M. I. and R. F. Mitchell. (1967) Chronometric analysis of classification. Psychol. Rev. 74, 392-409.

Studdert-Kennedy, M., A. M. Liberman, K. S. Harris, and F. S. Cooper. (1970) The motor theory of speech perception: A reply to Lane's critical review. Psychol. Rev. 77, 234-249.

Studdert-Kennedy, M., A. M. Liberman, and K. N. Stevens. (1963) Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. J. Acoust. Soc. Amer. 35, 1900.

The Role of Auditory Short-Term Memory in Vowel Perception*

David B. Pisoni+

ABSTRACT

The distinction between categorical and continuous modes of speech perception has played an important role in recent theoretical accounts of the speech perception process. Certain classes of speech sounds such as stop consonants are usually perceived in a categorical or phonetic mode. Listeners can discriminate between two sounds only to the extent that they have identified those stimuli as different phonetic segments. Recently, several findings have suggested that vowels, which are usually perceived in a continuous mode, may also be perceived in a categorical-like mode, although this outcome may be dependent upon various experimental manipulations. This paper reports three experiments that examined the role of auditory short-term memory in the discrimination of brief, 50 msec vowels and longer, 300 msec vowels. Although vowels may be perceived in a categorical-like mode, differences still exist in perception between stop consonants and steady-state vowels. The findings are discussed with regard to auditory and phonetic coding in short-term memory.

A basic assumption underlying recent theoretical work in speech perception has been that the perception of speech sounds involves processes and mechanisms that are somehow basically different from the processes involved in the perception of other auditory stimuli. One line of evidence cited in support of this view concerns the identification and discrimination of various classes of synthetic speech sounds (see Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). Some classes of speech sounds, such as stop consonants, have been found

---

*A short report of some of these findings was presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., April 1973.

+Indiana University, Bloomington.

Acknowledgment: This research was supported in part by a PHS Biomedical Sciences Grant (S05 RR 7031) to Indiana University and in part by a grant from NICHD to Haskins Laboratories. I am grateful to Michael Studdert-Kennedy, A. M. Liberman, and R. M. Shiffrin for their advice on this project. I would also like to thank my able assistants: D. L. Glanzman, J. Tash, J. Sawusch, and C. Lewis for their help in running experiments and analyzing data.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

to be perceived in a categorical mode; listeners can discriminate between two acoustically different stop consonants only to the extent that they can identify the stimuli as different on an absolute basis (Liberman, Harris, Hoffman, and Griffith, 1957; Liberman, Harris, Kinney, and Lane, 1961; Mattingly, Liberman, Syrdal, and Halwes, 1971; Pisoni, 1971). In contrast, other classes of speech sounds such as steady-state vowels have been found to be perceived in a continuous mode; listeners can discriminate among many more vowels than would be expected on the basis of absolute identification alone (Fry, Abramson, Eimas, and Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, and Ohman, 1969; Pisoni, 1971).

Differences in perception between stop consonants and steady-state vowels have not only played a central role in theoretical accounts of speech perception (see Liberman et al., 1967; Liberman, Mattingly, and Turvey, 1972; Studdert-Kennedy, 1973), but have also been implicated in several recent studies dealing with immediate recall of these two classes of speech sounds. For example, Crowder (1971, 1973a, 1973b) has reported that for lists of stop-consonant vowel syllables presented auditorily, a recency effect is observed in immediate serial recall if the syllables in the list contrast only on vowels (e.g., /bi/, /ba/, /bu/); however, the recency effect is curiously absent if the syllables contrast only on the stop consonants (e.g., /ba/, /da/, /ga/). The recency effect describes an advantage in recall of the last serial position over the second-to-last serial position in a list of items.

Crowder (1971, 1973a, 1973b) also reports two other differences in immediate memory for stop-consonant vowel stimuli: a modality effect and a suffix effect. The modality effect refers to the advantage of auditory over visual presentation for recall of items from later serial positions of a list. This modality effect has been observed for vowels but not for consonants. On the other hand, the suffix effect refers to a decrease in performance for items at the end of a list when a redundant word is presented after the last item in that list. The suffix effect has also been found with vowels but not with stop consonants. All three findings--the recency effect, the modality effect, and the suffix effect--are characteristic of a form of auditory memory called precategorical acoustic storage (PAS) by Crowder and Morton (1969). They argue that this form of memory holds some relatively unanalyzed representation of an acoustic stimulus for approxinately 2 sec. This form of "sensory" memory has been discussed recently by Massaro (1972a, 1972b) and it should be distinguished from his preperceptual auditory store which holds acoustic information for a much shorter period of tine (i.e., 250 msec) and which has distinctly different properties.

In the original study by Crowder (1971), information about the vowel and consonant was confounded by their position within the syllables. The stop consonants were in initial position in the syllable and the vowels were in final position. Similar results, however, have been reported by Cole (1973), who found that consonants show less of a recency effect than vowels, regardless of the initial or final position in the syllables of the critical to-be-remembered information (e.g., /ba/ vs. /ab/). Crowder (1973a) recently replicated these results in a study which controlled for position of the information within the syllable. Both Crowder (1971, 1973a) and Cole (1973) have explained the differences in recency effects for consonants and vowels in terms of differences in auditory memory for these two classes of speech sounds. Thus, they assume that

the recency effect for the vowels is due to retrieval of some auditory representation for vowels from a sensory memory store such as Crowder and Morton's PAS system. Crowder (1971, 1973a, 1973b) has been somewhat more specific and further suggests that auditory information in vowels, but not in stop consonants, is represented in PAS.

Crowder (1971, 1973a, 1973b), Liberman et al. (1972), and Cole (1973) have all noted the parallel between the differences in perception of stop consonants and vowels (the categorical vs. continuous distinction) and the differences in serial recall for these two types of stimulus vocabularies. These investigators have suggested that the differences in immediate recall may in fact be due to differences in perceptual processing for these two classes of speech sounds. For example, in discussing these results Liberman et al. (1972) state:

> ... the difference in recency effect between the stops and vowels is exactly what we would expect.... the special process that decodes the stops strips away all auditory information and presents to immediate perception a categorical linguistic event the listener can be aware of only as (b,d, g,p,t,k). Thus, there is for these segments no auditory, precategorical form that is available to consciousness for a time long enough to produce a recency effect. The relatively unencoded vowels, on the other hand, are capable of being perceived in a different way. Perception is more nearly continuous than categorical.... the auditory characteristics of the signal can be preserved for a while (p. 329).

The position described by Liberman et al. (1972) appears to be a reasonable explanation of the results showing differences in serial recall between consonants and vowels. We take these findings as being generally consistent with the assumption that the differences are perceptual in nature, presumably occurring at a relatively early stage of perceptual analysis. However, many of the perceptual studies dealing with the identification and discrimination of consonants and vowels have not been very specific about where the differences between these two classes of sounds occur during perceptual processing. In addition, although one might want to argue that the recall findings are due to differences in perceptual processing for consonants and vowels, some recent findings seem to indicate that vowels may also be perceived categorically, much like stop consonants. If vowels are perceived categorically in the same way and by the same mechanisms as stop consonants, we are clearly faced with somewhat of a dilemma in trying to account for the serial recall data by reference back to the perceptual findings. One way to deal with this problem would be to demonstrate that the categorical perception findings for the vowels are both qualitatively and quantitatively different from those obtained for the stop consonants.

In several previous reports, Pisoni (1971, 1973) has suggested that the major differences in discrimination between stop consonants and steady-state vowels are to be found in an examination of within-phonetic category comparisons. Discrimination performance for the putative categorically perceived vowels is well above chance within phonetic categories, suggesting an auditory as well as phonetic basis for a discrimination decision. The situation for the stop consonants is quite different. Under identical experimental conditions, subjects apparently

cannot retrieve the auditory information needed for a within-phonetic category decision with the consonants (Pisoni, 1973). This paper describes a revised model of the perceptual processes involved in the ABX test based on Fujisaki and Kawashima (1970) and then reports a series of experiments dealing with the discrimination of steady-state vowels. The major purpose of these studies was to make explicit some of the differences between the type of categorical-like perception recently observed with vowels and the type of categorical perception typically observed with stop consonants.

Auditory and Phonetic Memory Codes

Since Fujisaki and Kawashima's (1970) findings on categorical-like perception of vowels are central to a number of theoretical efforts in speech perception (Pisoni, 1973; Studdert-Kennedy, 1973), we consider some of their results and a modified version of their original model of the perceptual processes involved in the ABX discrimination test.

Fujisaki and Kawashima (1968, 1969, 1970) proposed a two-stage model of categorical perception, a model based on a distinction between auditory and phonetic information in short-term memory (STM). The model assumes that differences in discrimination between classes of speech sounds are due to the degree to which auditory and phonetic information is employed in the decision process in discrimination. Although not explicitly described by Fujisaki and Kawashima, we assume, following Studdert-Kennedy (1973), that auditory information is coded in short-term memory subsequent to an analysis of the acoustic waveform into a set of time-varying psychological dimensions such as pitch, loudness, and timbre. Similarly, we assume that phonetic information is coded as abstract phonetic features in STM after the "auditory" dimensions have made contact with some type of representation generated from synthesis rules residing in long-term memory (LTM).

The basic model proposed by Fujisaki and Kawashima (1969, 1970) is shown with several additions and modifications[1] in Figure 1. As shown, the model applies to discrimination exclusively within the ABX discrimination format, but the same assumptions could be adapted to other discrimination procedures.

---

[1] We assume that the encoding of speech sounds involves information about both the phonetic features of the stimulus and the auditory properties of the acoustic input. Furthermore, auditory information at relatively early stages of processing may be lost more rapidly from STS than the higher order phonetic information. According to this view, both an auditory and a phonetic representation are present in STS; the comparison process in ABX discrimination entails the retrieval of either the auditory trace or the phonetic code for a correct decision. One consequence of this view would be that if a S must base a decision on auditory information, he would be more likely to show better performance in an ABB triad than an ABA triad. This should be true for two reasons. First, the two stimuli in the ABB triad are closer together in time. Second, there is no interfering acoustic event in the comparison (retroactive) interval as is the case with the ABA triad. Glanzman and Pisoni (1973) examined these two types of comparisons in their data and found exactly these results for both stop consonant ABX discrimination (along the VOT continuum) and vowel ABX discrimination. Crowder (1973b) has recently alluded to this same observation.

Figure 1

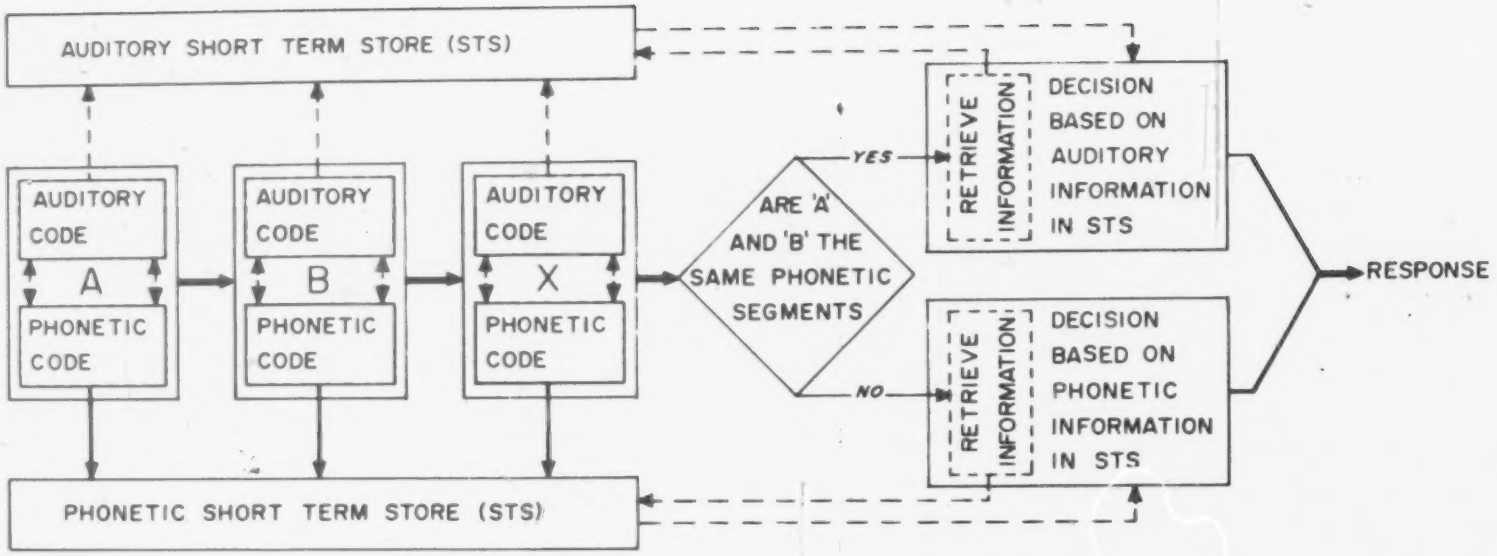


Figure 1: A schematic representation of the perceptual and decision processes in an ABX discrimination task. Test stimuli are presented and two parallel representations are stored in STS, an auditory code and a phonetic code. If the first two stimuli, A and B, have been recognized and encoded as different phonetic segments, then the comparison of stimulus X with A or B is based on phonetic information in STS. Otherwise, the comparison of X with A or B is based on auditory information in STS. Auditory information in STS is thought to be lost much faster than phonetic information.

According to this model, when a listener is required to discriminate between two different phonetic types, the decision in the discrimination task is based on phonetic information coded in STM. These derived phonetic properties or features of the auditory stimuli reside in phonetic short-term store. In this case, the listener determines whether the first two stimuli (i.e., A and B) are different phonetic segments. Since A and B are different phonetic segments, the listener's decision about X is based exclusively on a comparison of the phonetic information coded in STM. Thus, he compares X with B and X with A and then determines which is the closest match.

However, the situation is somewhat different when the listener is required to discriminate between two identical phonetic types; that is, two that are acoustically different but that have been drawn from the same phonetic category. Now the listener must rely exclusively on the auditory information for each stimulus coded in STM. In order to arrive at a correct decision in the discrimination task, the listener must retrieve and compare with stimulus X the auditory representations of the two stimuli in auditory short-term store, since the two stimuli, A and B, were not originally identified as different phonetic segments. The listener must make a comparative judgment based on auditory information of the acoustic properties of these stimuli rather than an absolute judgment based on the phonetic features.

The basic model first developed by Fujisaki and Kawashima (1969, 1970) and expanded here predicts that categorical perception is related to the degree to which auditory and phonetic information in STM can be employed in the decision process during ABX discrimination. It has been reported that the major differences in discrimination between stop consonants and steady-state vowels appear to be related to differences in retrieval of auditory rather than phonetic information from STM (Fujisaki and Kawashima, 1970; Pisoni, 1971, 1973; Pisoni and Lazarus, 1973). But the extent to which auditory and phonetic information is encoded and later retrieved from STM will depend on a number of factors--for example, the duration of the critical information in the signal; the acoustic environment or context of the cues; whether the acoustic cues are steady-state or transient; and the particular information-processing task. All these factors should presumably influence the way auditory and phonetic information is used by the decision rules in discrimination.

The experiments reported in this paper are concerned with three related questions about vowel discrimination and the role of auditory STM in speech perception. First, what effect does duration play in vowel discrimination? Fujisaki and Kawashima (1970) found that isolated steady-state vowels of very brief duration (50 msec) tend to be perceived in a categorical-like mode; there was a peak across the phonetic boundary and a trough within phonetic categories. However, although Fujisaki and Kawashima showed that perception of vowels was more nearly categorical at short durations, they did not employ stimuli with durations comparable to the earlier vowel perception studies conducted at Haskins Laboratories (Fry et al., 1962; Stevens et al., 1969). It is possible that the longer vowels of 300 msec also may be perceived in a somewhat categorical-like mode.

The second question deals with the effect of context. What role does the immediately surrounding acoustic environment have on vowel discrimination? Stop consonants always occur in syllabic context. Moreover, there is a relatively

complex relation between perceived phonetic segment and its representation in the acoustic signal; the essential acoustic cue for the stop consonants is a rapidly changing spectrum (F_1 and F_2 transitions) which is both short in duration (50 msec) and transient in nature (Liberman, Delattre, Cooper, and Gerstman, 1954). In contrast, the major acoustic cue for vowels, the frequencies of the first three formants, has a relatively long duration and is relatively uniform over the entire length of the stimulus (Delattre, Liberman, Cooper, and Gerstman, 1952). Stevens (1968), Sachs (1969), and Fujisaki and Kawashima (1969, 1970) have all found that vowels tend to be perceived more categorically when they appear in a fixed context than when the same stimuli are presented in isolation. Fujisaki and Kawashima suggested that the context served as a "perceptual reference" or anchor. However, it could be argued that the context selectively interfered with retention of the auditory information in target vowels. If this interference hypothesis is correct, vowel discrimination should be poorer when a reference context follows a target vowel (retroactive interference) than when it precedes it (proactive interference). We assume that the retroactive context acts to interrupt the processing of the target vowel as well as more generally to mask transitional information in a stop-consonant vowel syllable. In addition, the amount of interference should be related to the similarity of the target sound and context. For example, vowels should suffer more interference from other vowels than from tones or white noise (Darwin, 1971).

Finally, the third question deals with the ABX discrimination test that Fujisaki and Kawashima (1969, 1970) employed in their experiments on vowels. Is the categorical-like discrimination found with vowels in the ABX test also found more generally with other discrimination procedures (Pisoni, 1971)? We consider discrimination performance with vowels in another test procedure, the 4IAX test of a paired similarity, which was introduced in another report (Pisoni, 1971). If there are large differences between the ABX and the 4IAX test for both short, 50 msec vowels, and longer, 300 msec vowels, we will have additional evidence for suspecting that the type of categorical perception observed for vowels is somehow both qualitatively and quantitatively different from that observed for the stop consonants.

EXPERIMENT I

In this experiment we compare discrimination of short (50 msec) vowels with longer (300 msec) vowels. The major aim of the study was to replicate and extend the findings of Fujisaki and Kawashima (1970) and Pisoni (1971), who reported that vowels are perceived as more nearly categorical at short durations.

Method

Materials

Stimuli. Two sets of seven steady-state vowels were synthesized on the vocal tract analogue synthesizer at the Research Laboratory of Electronics, Massachusetts Institute of Technology. Table 1 provides the frequencies of the first three formants for both sets of vowels. The fourth and fifth formants were fixed at 3500 Hz and 4500 Hz respectively. The seven stimuli were arranged so that the first three formants varied in equal logarithmic steps from the English vowels /i/ through /I/. The formant frequencies chosen were identical to those used by Stevens et al. (1969) in their cross-language study of vowel perception.

TABLE 1: Formant Frequencies for vowel stimuli.

Stimulus Number	Formant Frequency (Hz)		
	F_1	F_2	F_3
1	270	2300	3019
2	285	2262	2960
3	298	2226	2902
4	315	2180	2836
5	336	2144	2776
6	353	2103	2719
7	374	2070	2666

One set of stimuli had a steady-state duration of 300 msec (the equivalent of approximately 30 glottal pulses) with a rise and decay time of 50 msec. The second set of stimuli had a steady-state duration of 50 msec (the equivalent of five glottal pulses) with a rise and decay time of 10 msec. Both sets of stimuli had identical formant frequency values and had a falling fundamental frequency. F_0 fell linearly from 125 Hz to 80 Hz for the long vowels and from 125 Hz to 100 Hz for the short vowels. Bandwidths of the first three formants were fixed at 50 Hz, 80 Hz, and 110 Hz respectively. The stimuli were originally recorded on magnetic tape at MIT and then digitized on the PCM system at Haskins Laboratories where the waveforms were stored on disc for test preparation. These stimuli are identical to those used by Pisoni (1971).

Experimental tapes. All the experimental tapes were produced under computer control from the digital values of these stimuli. A 1000 Hz tone was placed at the beginning of each tape to insure that the playback levels would be uniform throughout the testing sessions.

Four different 70-item identification test sequences were prepared for each set of vowel stimuli. Each identification test contained ten different randomizations of an entire set of seven stimuli. The stimuli were recorded singly with a 4 sec interval between presentations and an 8 sec interval after every ten trials.

Four different 88-trial ABX discrimination tapes were also constructed for each set of stimuli. All possible one- and two-step comparisons of the seven stimuli appeared twice within each tape. The tapes were balanced, with the restriction that each ABX triad occur equally often within each half of the test. Stimuli within each triad were separated by 1 sec, while successive triads were separated by 4 sec. There was an 8 sec pause after every ten trials.

Subjects

Eighteen undergraduate students at Indiana University served as Ss. They were obtained from the Psychology Department's subject pool and received either 1 hour course credit or $1.50 for each session. All Ss were right-handed native speakers of English with no history of a hearing or speech disorder. None of the Ss had heard any synthetic speech before the present experiment.

Procedure

The experimental tapes were reproduced on a high quality tape recorder (Ampex AG-500) and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder playback was adjusted to give a voltage across the earphones equivalent to 70 db SPL re 0.0002 dyn/cm^2 for a 1000 Hz calibration tone. To compensate for the differences in loudness between the 300 msec and 50 msec vowels due to stimulus duration, the gain for the calibration tone on the 50 msec vowel tapes was adjusted by means of decade attenuators to be +8 db above the 300 msec vowels. Measurements were made on a VTVM (Hewlett-Packard Model 1051) before presentation of each experimental tape. Ss were run in two counterbalanced groups of nine Ss each. They were tested in a small experimental classroom. All Ss in a given session heard the same stimuli in the same order.

The E read aloud to the Ss a set of instructions which explained the nature of the experiment. Ss also had a set of printed instructions before them. Ss were told that this was an experiment dealing with speech perception. For the identification tests, Ss were required to identify each stimulus as either the vowel /i/ as in "beet" or /I/ as in "bit." In the ABX discrimination tests Ss were told that the stimuli would be arranged in groups of three and that their task was to decide whether the third sound was more like the first sound or the second sound. Ss were told to guess if they were not sure, but to respond on every trial. Judgments were recorded in prepared response booklets.

Ss were run for an hour a day on four consecutive days. On the first two days one group received the 300 msec vowels and the other group received the 50 msec vowels. The conditions were reversed for each group on the last two days. An identification test for a given stimulus condition was always followed immediately by the corresponding ABX discrimination tests. When the data are combined over the four sessions, each S provided 40 identification responses to each of the 7 stimuli in both the 300 and 50 msec vowel conditions. Each S also provided 32 judgments for each of the AB discrimination comparisons in each stimulus condition.

Results

The probabilities of identification averaged over the 18 Ss for each stimulus condition are given in Table 2. The identification probabilities for both stimulus conditions are quite sharp and consistent. Examination of Table 2 shows that the probabilities of identification for the two vowel conditions are very nearly exact complements of each other. There is a slight shift in cross-over point or phonetic boundary between /i/ and /I/ as stimulus duration is reduced from 300 msec to 50 msec; the boundary shifts predictably in favor of the short, lax vowel /I/.

TABLE 2: Probabilities of identification averaged over 18 Ss for 300 msec and 50 msec stimuli.

	300 msec Vowels						
	Stimulus Number						
	1	2	3	4	5	6	7
/i/	.998	.997	.968	.635	.141	.033	.018
/I/	.002	.003	.032	.365	.859	.967	.982

	50 msec Vowels						
	Stimulus Number						
	1	2	3	4	5	6	7
/i/	.980	.979	.871	.419	.071	.025	.021
/I/	.020	.021	.129	.581	.929	.975	.979

The average one- and two-step obtained ABX discrimination functions are shown in Figure 2 for both stimulus conditions. The predicted discrimination functions, which were derived from the identification probabilities according to the Haskins Laboratories' model (Liberman et al., 1957; Pollack and Pisoni, 1971) are also plotted in Figure 2. The predicted functions represent what would be expected under the strong categorical perception assumption: that discrimination is no better than absolute identification.

The obtained discrimination functions for both vowel conditions show peaks at the phonetic boundary and troughs within phonetic categories. Analysis of variance indicates that discrimination performance is significantly better on the 300 msec vowels than on the 50 msec vowels, $F(1,16) = 9.59$, $p < .01$, but only for the one-step comparisons. This finding is consistent with Fujisaki and Kawashima (1970) and Pisoni (1971). The two-step obtained discrimination functions did not differ significantly from each other. There was a significant difference between obtained and predicted discrimination scores for both the one- and two-step comparisons, $F(1,16) = 77.27$, $p < .001$ and $F(1,16) = 343.12$, $p < .001$, respectively.

We may obtain a better quantitative idea of these results by comparing the obtained discrimination functions to those predicted from the model of categorical perception. We assume that in the ideal case of categorical perception there will be an exact mapping of the discrimination functions predicted from identification and the functions obtained in ABX discrimination. Although an exact mapping is rarely found, since the obtained functions are usually higher than the predicted, we can use this assumption to our advantage for comparative purposes.

Figure 2

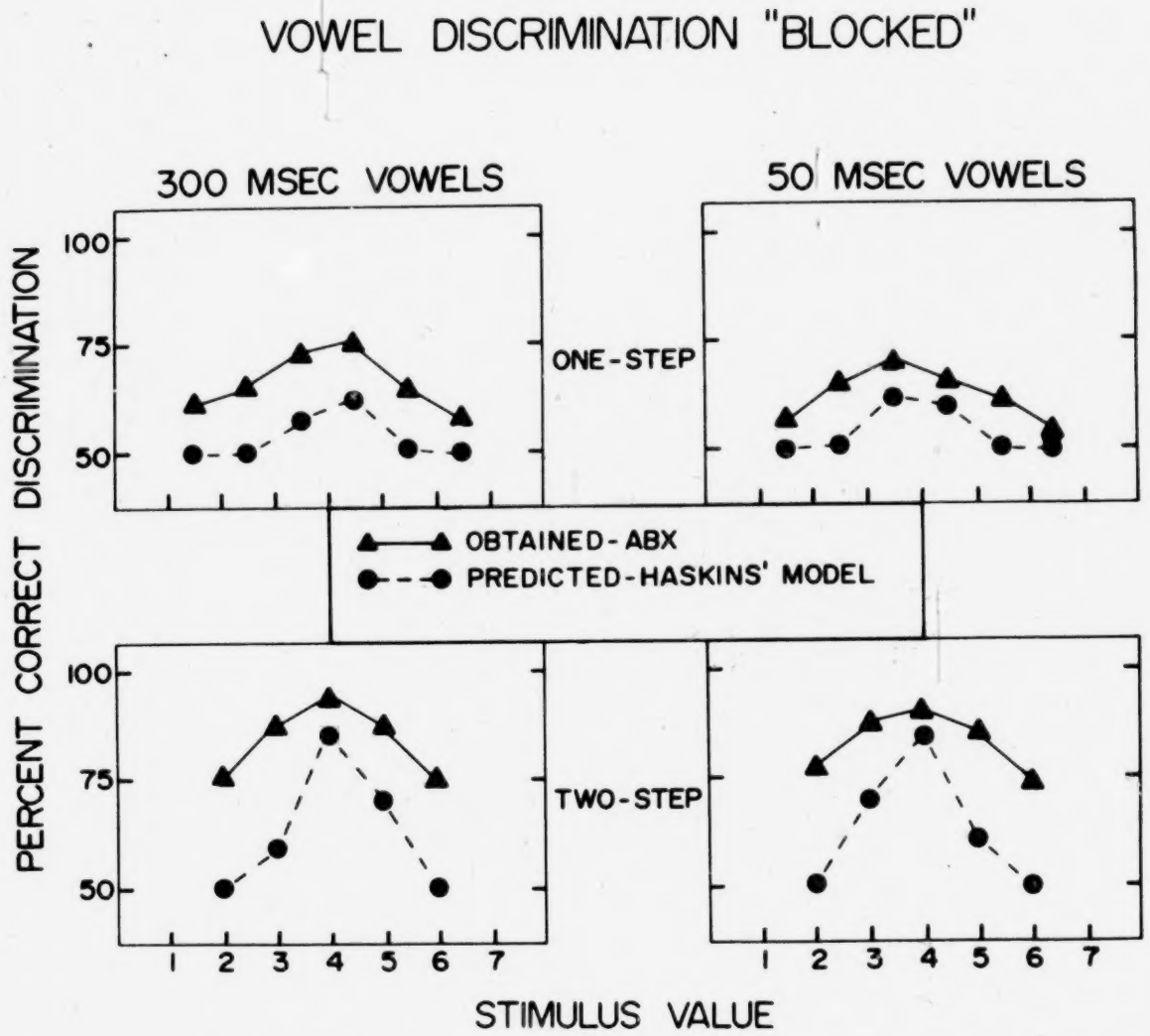


Figure 2: ABX vowel discrimination functions for long (300 msec) and short (50 msec) stimuli averaged over 18 Ss under blocked presentation. The dashed lines show the predicted discrimination functions derived from the Haskins' model of categorical perception.

We assume that the difference between the obtained and predicted discrimination functions for a given condition represents a measure of the degree to which that particular condition deviates from the predictions of the idealized model. Hence, it follows that the smaller the discrepancy between the obtained and predicted functions, the closer the obtained discrimination function will be to the categorical model. If short vowels are more categorical than longer vowels, we would expect a smaller difference between the obtained and predicted functions for the 50 msec condition than for the 300 msec condition. The analyses reported below were carried out by first calculating difference scores on the obtained and predicted data for each S and then performing separate analyses of variance on the one- and two-step scores. Of greatest interest is the comparison between the long and short vowel conditions.

Analysis of variance on the one-step difference scores revealed a significant effect for stimulus duration, $F(1,16) = 12.80$, $p < .005$. The difference between the obtained and predicted scores was greater for the longer vowels than for the shorter vowels. There was also a significant main effect for stimulus comparison, $F(5,80) = 3.68$, $p < .01$. None of the interactions reached statistical significance.

A similar analysis on the two-step data failed to find a significant difference for the main effect of vowel duration, although the stimulus comparison did reach significance again, $F(4,64) = 9.44$, $p < .001$. In addition, the vowel duration by stimulus comparison interaction was significant, $F(4,64) = 3.79$, $p < .01$.

Discussion

The results of this experiment seem to indicate that vowels of both long and short duration may be perceived in a categorical-like mode. Differences in discrimination, as they are related to stimulus duration, are revealed only in the one-step comparisons. This finding is consistent with the results reported by Fujisaki and Kawashima (1970) and Pisoni (1971). Although there was no overall effect of vowel duration for the two-step data, differences restricted to particular types of stimulus comparisons along the continuum did occur. These results appear to be due to the apparent difference in the location of the phonetic boundary between /i/ and /I/ under the two conditions of duration. Since Fujisaki and Kawashima (1970) employed only one-step stimulus comparisons, the present two-step data have little bearing on their results or conclusions in this regard.

The major outcome of this experiment may appear to be somewhat at variance with previous studies of vowel discrimination, particularly the original vowel perception studies conducted by investigators at Haskins Laboratories (Fry et al., 1962; Stevens et al., 1969). In these studies, vowel discrimination was described as more nearly continuous than categorical. However, Stevens et al. (1969) did find some evidence for peaks in the discrimination functions which were correlated with changes in identification, but the troughs in the discrimination functions were well above chance when contrasted with the discrimination data typically found with the stop consonants. Although the discrimination functions, particularly the two-step data, appear by inspection to be categorical, we note that performance within phonetic categories is in fact well above chance. An auditory, nonphonetic basis for discrimination is available to the listener.

One of the major weaknesses of the original Haskins' model of categorical perception is its failure to account for within-category discrimination performance that may be at a level well above chance. In the model (Liberman et al., 1957; Liberman et al., 1961; Pollack and Pisoni, 1971) it was assumed explicitly that if a listener identifies two stimuli as the same he can discriminate them only by chance.

In the model developed by Fujisaki and Kawashima (1970), performance that is above chance on within-category comparisons is assumed to reflect the underlying contribution of auditory short-term memory to ABX discrimination. Thus, Fujisaki and Kawashima do not assume that discrimination is at chance within categories, but rather that the level of within-category performance reflects the contribution of auditory short-term memory to discrimination.

Two assumptions are implicit in the model shown in Figure 1. First, discrimination will be based on phonetic information if the first two members of an ABX triad (A and B) are judged by the listener to be <u>different</u> phonetic segments. Second, discrimination will be based on auditory short-term memory if the first two members of an ABX triad have been judged to be the <u>same</u> phonetic segments.

Following Fujisaki and Kawashima, the predicted correct ABX discrimination score may be expressed by the following two components:

$$C_{ABX} = C_{A \neq B} + C_{A=B}$$
$$= C_{A \neq B} + M_{A=B} \cdot P_{A=B}$$

where $C_{A \neq B}$ = the probability that a correct discrimination occurs on the basis of phonetic identification.

$C_{A=B}$ = the probability that a correct discrimination occurs on the basis of auditory short-term memory.

$P_{A=B}$ = the probability that stimuli A and B are identified as the same phonetic segments.

$M_{A=B}$ = the conditional probability that a correct discrimination takes place when A and B are identified as the same phonetic segments. This quantity indicates the degree to which judgments are based on auditory short-term memory and is equal to the asymptotic value of C_{ABX} at the extremes of the stimulus range (i.e., within-category comparisons).

These components are related according to the following equations:

$$C_{A \neq B} = 1/2 \left[(P_1 - P_2) + P_1(1 - P_2) + P_2(1 - P_1)\right]$$

$$C_{A=B} = \left[P_1P_2 + (1 - P_1)\ (1 - P_2)\right] \cdot M_{A=B}$$

P_1 and P_2 represent the probabilities that stimuli A and B in the triad are identified as the same phonetic segments in an absolute identification test.

A new set of predicted ABX discrimination scores was obtained from the model outlined above. Figure 3 shows the obtained one- and two-step discrimination functions along with the new predicted functions derived from Fujisaki and

Figure 3

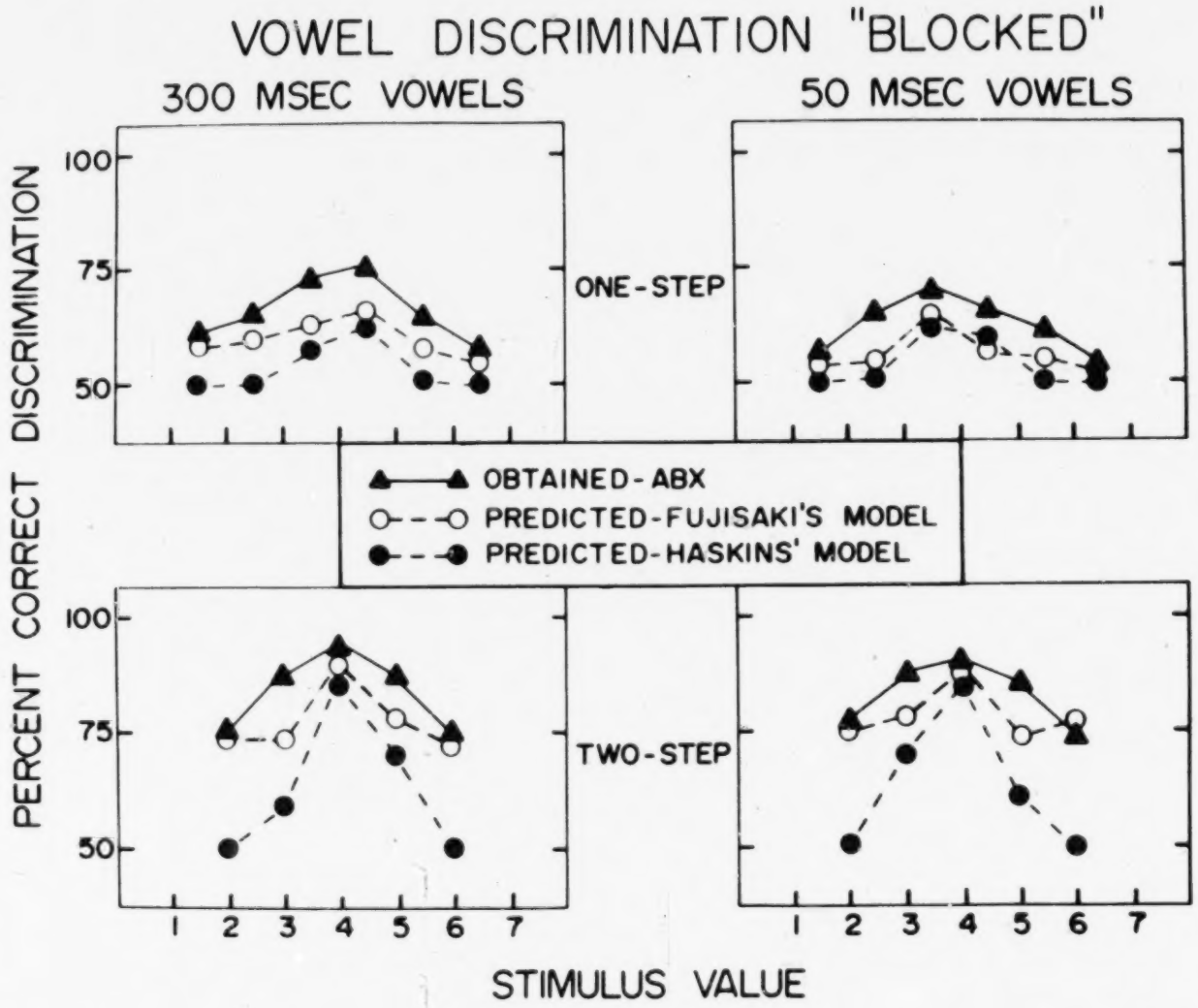


Figure 3: ABX vowel discrimination functions along with the predicted functions derived from Fujisaki's model. The Haskins' functions are also plotted for comparison.

Kawashima's model. Examination of this figure indicates that the new predicted discrimination functions match the obtained functions much more closely than the traditional Haskins' predictions. However, it should be pointed out that the better fit of the obtained data is to be expected since one parameter from the obtained data has been used in the predictions. The simplicity and advantage of the Haskins' model lies in the fact that no additional assumptions or data are required to predict discrimination performance under the strong categorical assumption.

It is possible that the results of this experiment were somehow due to the particular test procedures used. Although there was no significant main effect for order of presentation nor any interactions in the raw score analysis of variance, there appeared to be some slight differences in discrimination of short vowels depending upon the order of presentation over the experimental sessions. Discrimination of the short vowels was better if these stimuli were presented on the last two days of the experiment than if they were presented on the first two days.

An additional experiment was run to examine the possibility that test order effects might be responsible for some of the differences. Seven new Ss were obtained and run in two completely randomized and counterbalanced groups; there were four Ss in one group and three in the other. The same experimental tapes and procedures were used. The only difference introduced was that Ss received both short and long vowels on each day of testing, with the order reversed for each group.

Figure 4 shows the one- and two-step discrimination functions for short and long vowels. These results are basically quite similar to those obtained in the main study. An analysis of variance on the differences between the obtained and predicted discrimination scores was performed. The results indicated that neither the main effects (i.e., vowel duration and stimulus comparison) nor any of the interactions were significant. Thus, the differences between long and short vowels for the one-step data found in the previous experiment do not occur when possible order effects are controlled across testing sessions.

Inspection of Figure 4 reveals that discrimination is also somewhat categorical. The one-step obtained functions seem to map onto the predicted Haskins' functions reasonably well, although they differ systematically by a constant. The two-step obtained discrimination functions are quite similar to those of the larger experiment. They also differ from the predicted scores.

Using Fujisaki and Kawashima's model, predicted discrimination functions were also calculated for these data. Figure 5 shows these new functions along with the Haskins' predictions. Again, a better fit is obtained by accounting for the contribution of auditory short-term memory to discrimination.

To summarize, these experiments have shown categorical-like discrimination functions for both short (50 msec) vowels and longer (300 msec) vowels. Although there were peaks in the discrimination functions, the level of within-category discrimination was well above chance expectation. When the contribution of auditory short-term memory is included in the predicted discrimination functions, according to Fujisaki and Kawashima's model, relatively better fits are obtained for the observed discrimination scores for both vowel conditions. These results

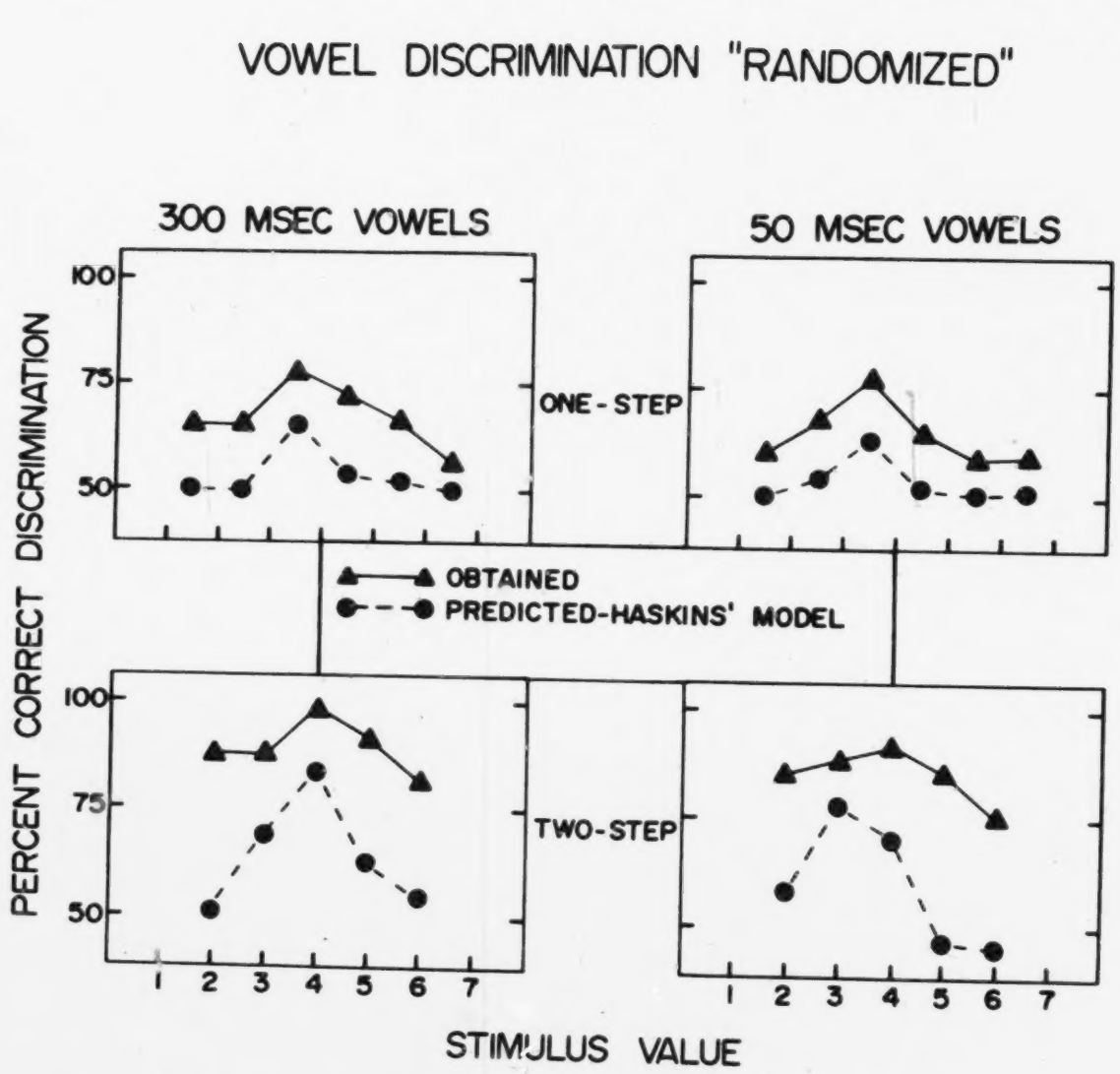


Figure 4

Figure 4: ABX vowel discrimination functions for long and short vowels for 7 Ss under randomized presentation. The dashed lines indicate the Haskins' predictions.

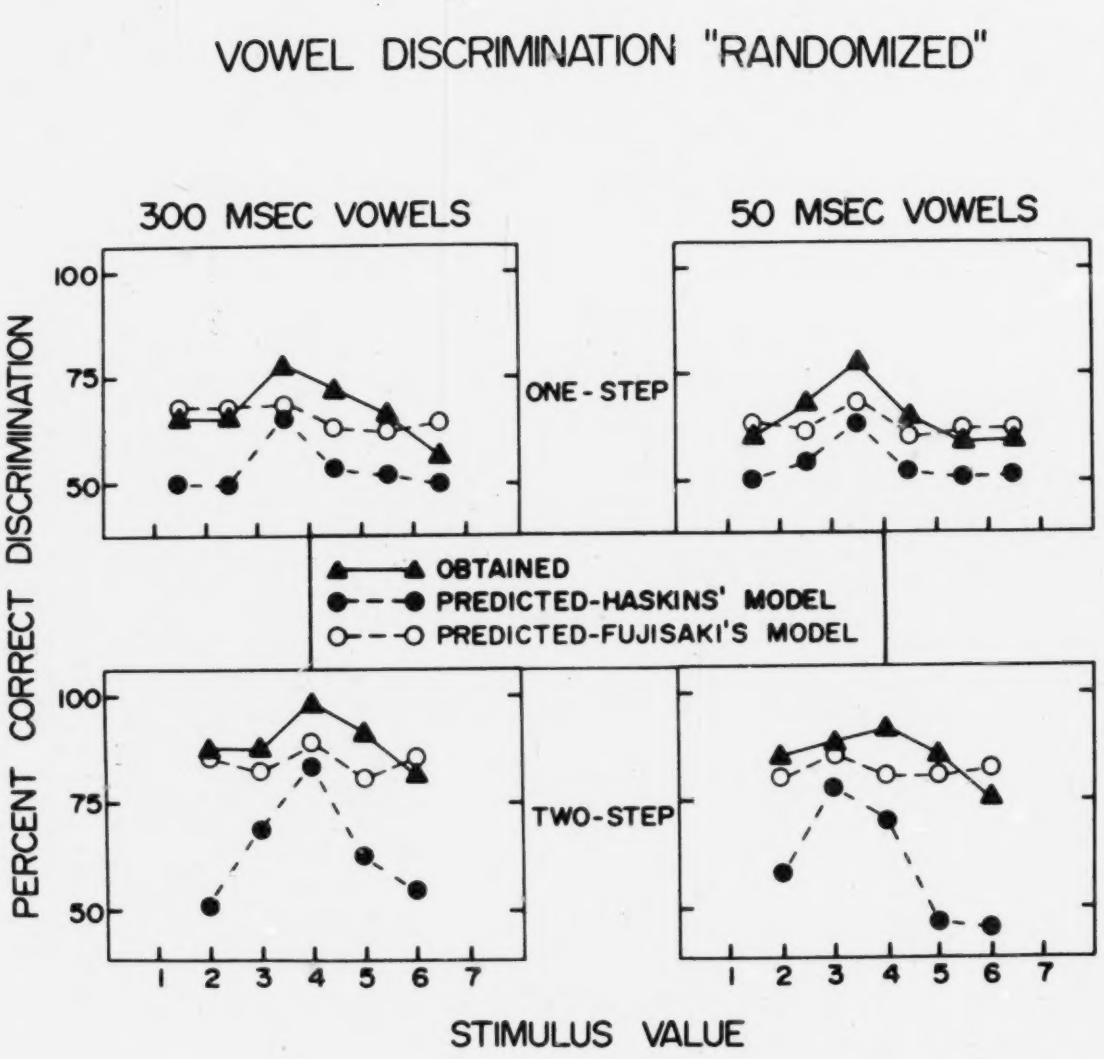


Figure 5

Figure 5: Randomized vowel discrimination functions with both the Fujisaki and Haskins' predicted functions.

suggest two conclusions. First, the role of stimulus duration taken alon appears to have relatively little effect on the shape of the discrimination functions. This is in agreement with an earlier observation reported by Pisoni (1971). Second, the type of categorical perception observed with these vowels is basically different from that observed in previous studies with the stop consonants. We suggest that the peaks and troughs in discrimination observed in the present study are primarily due to the nature of the ABX test procedure. The arrangement of stimuli in this test format may prevent listeners from retrieving the auditory information needed for discrimination and subsequently may force listeners to rely more heavily on phonetic coding in short-term memory.

EXPERIMENT II

This experiment is concerned with interference effects in vowel discrimination. Stevens (1968), Sachs (1969), and Fujisaki and Kawashima (1970) have reported that vowels in fixed contexts are perceived more categorically than the same vowels in isolation. Fujisaki and Kawashima (1970) suggested that the added context served as a "perceptual anchor" or reference. However, the context could act to interfere selectively with the retention of both auditory and phonetic information. If the perceptual anchor hypothesis is correct, it should be of little consequence where the reference context is placed (i.e., before or after the target vowel). However, if the context does selectively interfere with the encoding and retention of information, then temporal position of the reference or interference sound should show differential effects on discrimination. In addition, interference should be related to the similarity of the context and target vowels.

Method

Materials

Stimuli. The 50 msec short vowel continuum from the first experiment was used as the basic stimulus set. Four types of interfering stimuli were then constructed and used as contexts for each of the original seven stimuli. The interfering stimuli were 50 msec in duration and equal in overall intensity to the original vowels. The stimuli consisted of the following: (1) a 1000 Hz pure tone, (2) a burst of white noise, (3) the vowel /a/, and (4) the vowel /ɛ/. Each type of interfering context either preceded the target vowel (proactive interference condition) or followed the target vowel (retroactive interference condition). The original set of seven vowels was also presented alone as a control, and will be referred to as the silent condition.

Experimental tapes. Two types of identification and discrimination tests were prepared for each of the four types of interfering contexts. Two different 70-item identification tests were prepared for each of the proactive and retroactive interference conditions. In addition, four different 88-item ABX discrimination tapes were also constructed for each of these conditions. The identification and discrimination tapes for the silent condition were the same as those used in the previous two experiments. The test orders and timing sequences paralleled the test orders described in Experiment I.

Subjects

Twenty undergraduate students served as Ss. They were paid at the rate of $2.00 per hour for their services and met the same requirements as those Ss used in the previous experiments.

Procedure

The procedure was similar to that used in the previous experiment except for the following differences. Ss were run in four separate groups of five Ss each. Each group received one of the four interference conditions. Under a given condition, Ss received identification and discrimination tests for the silent (control) condition, and for the proactive and retroactive interference conditions.

The instructions were the same as those used previously, except that when Ss were run under proactive or retroactive interference conditions they were told to ignore the interfering sound and to try to concentrate on only the target vowels, /i/ and /I/.

Ss were run for one and a half hours a day on two consecutive days. On each day Ss received the silent vowel condition first, followed then by the proactive and retroactive conditions in differing order. Before each ABX discrimination test, Ss received the corresponding identification test: silent, retroactive, or proactive condition.

Results and Discussion

Table 3 shows the average one- and two-step percent correct discrimination scores for the silent, proactive, and retroactive context conditions under each of the four types of interference. These scores have been summed over the stimulus comparisons. Discrimination performance is generally lowest in the retroactive condition and highest in the silent condition for each type of interference. Analyses of variance were performed separately on the one- and two-step scores. The main effect for context position was significant for both analyses: $F(2,24) = 5.98$, $p < .01$ for the one-step scores and $F(2,24) = 23.65$, $p < .001$ for the two-step scores. Although the main effect for type of interference did not reach significance in either analysis, there was a significant interaction between type of interference and context position for the two-step data, $F(6,24) = 3.76$, $p < .01$.

The major results of this study are predicted by the interference assumption: there is more retroactive interference than proactive interference in vowel discrimination. Moreover, as shown in Table 3, there is more retroactive interference for a more similar vowel (e.g., /ɛ/) than a less similar vowel (e.g., /a/) for targets /i/ and /I/. The interaction between context position and type of interference may also be due in part to the relatively better performance for the tone condition in all context positions. This result is not entirely surprising since we would expect tonal stimuli to have little effect on the initial encoding process for the target vowels used here.

To summarize, this study provides evidence for interference effects in the discrimination of vowels in the ABX test paradigm. These effects seem to be

TABLE 3: Average percent correct for context position for each of four types of interference.

	One-Step Discrimination Context Position			Two-Step Discrimination Context Position		
	Silent	Proactive	Retroactive	Silent	Proactive	Retroactive
White Noise	59	60	58	78	75	70
1000 Hz Tone	62	59	54	79	82	76
Vowel /a/	60	60	57	85	75	73
Vowel /ɛ/	64	53	55	78	71	59
Mean:	61	58	57	80	76	70

greater when the interfering context follows a target vowel than when it precedes the vowel. Moreover, the similarity of context and target vowel may be related to some interference with the initial encoding process when both auditory and phonetic information is registered in short-term store.

The results of this study have several implications. First, these findings argue against Fujisaki and Kawashima's (1970) general "perceptual anchor" hypothesis because they indicate relatively specific temporal relations and similarity effects in discrimination. Second, these results may be generalized to the case of stop consonant syllables. It is possible that the extended duration of the vowel in a stop consonant syllable may act as a backward masking stimulus and preserve the integrity of the syllable as a perceptual unit (Massaro, 1972a; McNeil and Repp, 1973).

EXPERIMENT III

In this experiment we compare vowels of both short and long duration under two discrimination procedures; the traditional ABX test and the 4IAX test of paired similarity. If the categorical-like discrimination observed with these vowels in Experiment I is due mainly to the nature of the ABX test, we should expect to find differences between these two types of discrimination procedures. Moreover, since the differences in vowel discrimination appear to be due primarily to the availability of auditory information, we anticipate advantages in discrimination to reveal themselves on within- rather than between- phonetic category comparisons.

Figure 6 shows the arrangement of stimuli in the traditional ABX test and the 4IAX test of paired similarity. In the ABX test, pairs of stimuli are arranged in triads; the first two stimuli are always different, the third stimulus is identical with either the first (A) or the second stimulus (B). This discrimination procedure requires that the subject encode and store each of the three stimuli over a relatively long time (e.g., several seconds) before arriving at a decision.

In the 4IAX test, two pairs of stimuli are presented on every trial; one pair is always the same and one pair is always different. The Ss' task is to determine which pair contains the same stimuli, the first pair or the second pair. We assume that the 4IAX is more sensitive to purely auditory information since a decision can be made on a pair-wise comparison. The first two stimuli are compared and a difference, d_1, is calculated and stored in short-term memory. The second pair of stimuli are compared and their difference, d_2, is also calculated and stored. A final decision may be obtained when the two differences are later recalled and compared.

Method

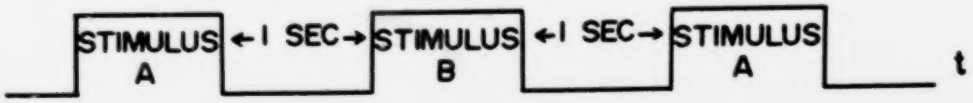
Materials

Stimuli. The 50 msec short vowel continuum and the 300 msec long vowel continuum from Experiment I were used.

Experimental tapes. The same identification tapes and ABX discrimination tapes from Experiment I were also used here. In addition, a new set of discrimination tapes were prepared in 4IAX format for both vowel conditions. All possible

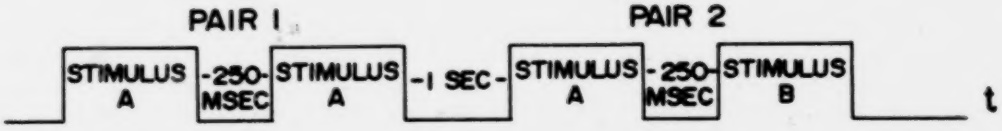
DISCRIMINATION TESTS

ABX TEST - PAIRS OF STIMULI ARRANGED IN TRIADS:
ABA, BAB, ABB, BAA



QUESTION: IS THE THIRD STIMULUS MORE LIKE THE FIRST OR SECOND STIMULUS ?
RESPONSE: FIRST OR SECOND STIMULUS

4IAX TEST - TWO PAIRS OF STIMULI ARE PRESENTED ON EACH TRIAL. ONE PAIR IS THE SAME AND ONE PAIR IS DIFFERENT: A-A—A-B, A-B—A-A, A-A—B-A, ETC.



QUESTION: WHICH PAIR WAS MORE SIMILAR - THE FIRST PAIR OR THE SECOND PAIR ?
RESPONSE: FIRST OR SECOND PAIR

Figure 6: Details of the two discrimination procedures; the standard ABX test and the 4IAX test of paired similarity.

one- and two-step comparisons of the seven stimuli in each continuum were employed and arranged in the following 4IAX sequences: AA-AB, AA-BA, AB-AA, and BA-AA. Four different 88-item discrimination tapes were produced under computer control. The stimuli within each pair were separated by 150 msec, and stimulus pairs were separated by one sec. Successive trials were separated by five sec. After every ten trials there was an extra ten-sec pause.

Subjects

Fourteen undergraduate students served as Ss. They were either paid for their services or received the equivalent in credit hours for their participation as part of a course requirement. They met the same requirements as the Ss used in the previous experiments.

Procedure

The fourteen Ss were run in two groups of seven Ss each. One group was assigned to the long (300 msec) vowel condition; the other group was assigned to the short (50 msec) vowel condition. Thus, vowel duration was a between-Ss variable and the discrimination test type was within-Ss variable.

On each day, Ss first received the standard identification test for a given vowel condition. This was followed by both types of discrimination tests. Four Ss in each group received the discrimination tests in one order while the other three Ss were presented with the reverse arrangement.

The instructions for identification and ABX discrimination were identical to those used in Experiment I. For the 4IAX discrimination test, Ss were told that they would hear two pairs of sound on each trial and that their task was to determine which pair sounded more similar, either the first pair or the second pair.

Results and Discussion

Table 4 shows the average probabilities of identification for each stimulus condition. The data are averaged over the 7 Ss in each vowel condition. These data are almost identical to the probabilities obtained in the first experiment.

Figure 7 shows the obtained discrimination functions for ABX and 4IAX discrimination for the two vowel conditions. Inspection of this figure reveals relatively large differences in discrimination between the two types of test procedures. Performance is much better at every stimulus comparison for the 4IAX test than for the ABX test. This is true for both vowel conditions, although the effects are most noticeable for the long, 300 msec vowels. The difference between the two discrimination tests was highly significant for both the one- and two-step comparisons, $F(1,12) = 36.10$, $p < .001$; and $F(1,12) = 21.85$, $p < .001$, respectively. The main effect of vowel duration did not reach significance in either the one- or two-step analysis.

The most interesting result, however, is the interaction between type of discrimination test and stimulus comparison along the continuum. Both the one-step and two-step interactions were significant, $F(5,60) = 3.79$, $p < .01$ and $F(4,48) = 3.76$, $p < .01$. This result, taken together with the main effect of

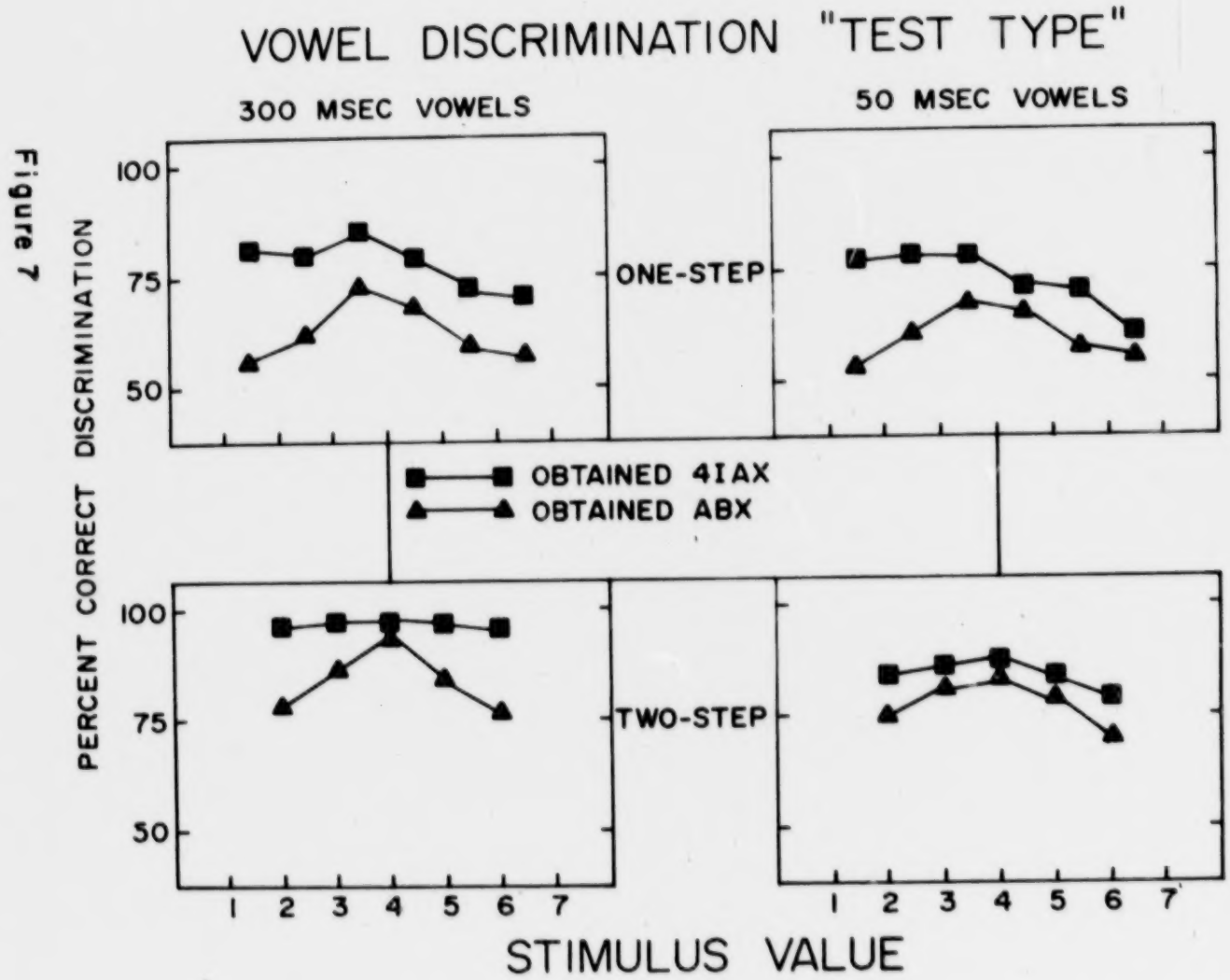


Figure 7

Figure 7: Average discrimination functions for long and short vowels under ABX and 4IAX test conditions. The functions are based on 7 Ss in each vowel duration condition.

TABLE 4: Probabilities of identification averaged over 14 Ss for 300 msec and 50 msec stimuli.

300 msec Vowels

	Stimulus Number						
	1	2	3	4	5	6	7
/i/	1.000	1.000	.967	.681	.157	.005	.010
/I/	.000	.000	.033	.319	.843	.995	.990

50 msec Vowels

	Stimulus Number						
	1	2	3	4	5	6	7
/i/	.967	.971	.824	.338	.119	.043	.029
/I/	.033	.029	.176	.662	.881	.957	.971

test type suggests that discrimination performance is not only better in the 4IAX test format but also that the shapes of the two discrimination functions are quite different. This result may be seen most clearly in the two-step discrimination functions for the 300 msec vowels. A very distinct advantage of the 4IAX test over the ABX test for within-phonetic category comparisons may be seen in this data. Discrimination in the 4IAX test may be thought of as more nearly continuous than categorical with these stimuli.

We conclude that the advantage in discrimination is due to the retrieval of auditory information. As noted earlier, the ABX test forces Ss to rely more extensively on phonetic rather than auditory coding in STM. Moreover, these results suggest that the categorical-like discrimination observed in Experiment I for both long and short vowels was probably due almost exclusively to the particular constraints of the ABX test rather than to some inherent property of the stimuli or to a limitation on the sensory capacities of the Ss.

More generally, it would appear that the form of categorical discrimination observed with the vowels is in fact different from that observed with the stop consonants. Comparable manipulations of the experimental procedures (i.e., use of the 4IAX test) with a stop consonant continuum have thus far failed to show equivalent changes in either the overall level or the shape of the discrimination functions (Pisoni, 1971). Figure 8, taken from a recent paper by Pisoni and Lazarus (1973), shows the results obtained under ABX and 4IAX discrimination with a synthetic consonant continuum ranging in voice onset time from /ba/ through /pa/. These consonant data were collected under the same experimental conditions as for the vowel data in the present study. Ss first took an absolute identification test and then received either an ABX test or a 4IAX test. The discrimination functions show some slight advantage in favor of the 4IAX test, but overall

Figure 8

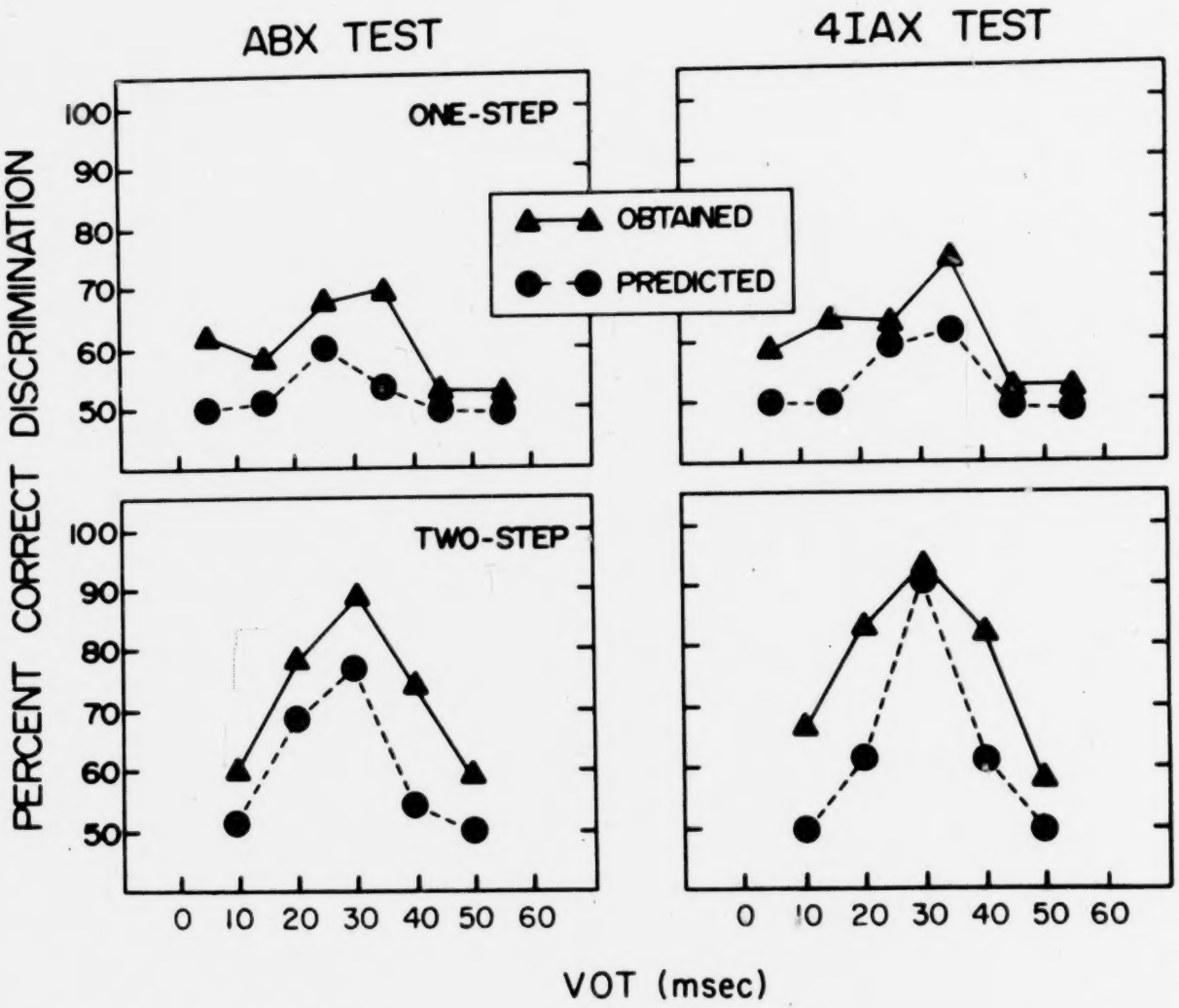


Figure 8: Average discrimination functions obtained with the ABX and 4IAX tests for a synthetic stop-consonant continuum varying in voice onset time from /ba/ to /pa/. Data are taken from Pisoni and Lazarus (1973).

the obtained functions still match the predicted ones fairly well. We do not preclude the possibility that auditory information can be employed in consonant discrimination; rather, we assume that auditory information from the earliest stages of processing tends to be lost from STM much more rapidly than phonetic information. As a result, decisions that require phonetic coding will be more accurate and reliable than decisions that require a comparison of auditory information in STM.

GENERAL DISCUSSION

The experiments reported in this paper have been concerned with the role of auditory short-term memory in vowel perception and more generally with the relationship between auditory and phonetic coding in speech perception. The main findings of these studies indicate that vowels of both short (50 msec) and longer (300 msec) duration appear to be discriminated in a categorical-like mode; there is a peak in the ABX discrimination functions for stimulus comparisons selected from different phonetic categories and a trough in these discrimination functions for comparisons selected from within the same phonetic category. The role of stimulus duration *per se* was shown to play a relatively minor role in contributing to the shape and level of the discrimination functions. The categorical-like discrimination for the vowels was assumed to reflect the greater dependence on phonetic rather than auditory coding in the ABX format. Support for this conclusion was obtained in two additional experiments. One study demonstrated specific temporal and similarity interference effects in ABX discrimination; the other study showed that vowel discrimination could be substantially improved when auditory information in STM is made more readily available for use in discrimination.

A major issue in speech perception has been the distinction between categorical and continuous modes of processing as reflected in the differences in discrimination between consonants and vowels. Despite several recent findings, we conclude that meaningful and theoretically important differences still exist between consonants and vowels. Moreover, we suggest that differences between categorical and continuous discrimination are primarily due to a failure of retrieval of auditory information in STS. The earliest stages of auditory processing of speech sounds tend to be lost from subsequent processing. Loss of this information may be due to both interference from succeeding acoustic events and to the decay of auditory information over time.

REFERENCES

Cole, R. A. (1973) Different memory functions for consonants and vowels. Cog. Psychol. 4, 39-54.

Crowder, R. G. (1971) The sound of vowels and consonants in immediate memory. J. Verbal Learn. Verbal Behav. 10, 587-596.

Crowder, R. G. (1973a) Representation of speech sounds in precategorical acoustic storage. J. Exp. Psychol. 98, 14-24.

Crowder, R. G. (1973b) Precategorical acoustic storage for vowels of short and long duration. Percept. Psychophys. 13, 502-506.

Crowder, R. G. and J. Morton. (1969) Precategorical acoustic storage (PAS). Percept. Psychophys. 5, 365-373.

Darwin, C. J. (1971) Dichotic backward masking of complex sounds. Quart. J. Exp. Psychol. 23, 386-392.

Delattre, P. C., A. M. Liberman, F. S. Cooper, and L. J. Gerstman. (1952) Observations on one- and two-formant vowels synthesized from spectrographic patterns. Word 8, 195-210.

Fry, C. B., A. S. Abramson, P. D. Eimas, and A. M. Liberman. (1962) The identification and discrimination of synthetic vowels. Lang. Speech 5, 171-189.

Fujisaki, H. and T. Kawashima. (1968) The influence of various factors on the identification and discrimination of synthetic speech sounds. Reports of the 6th International Congress on Acoustics, Tokyo, August.

Fujisaki, H. and T. Kawashima. (1969) On the modes and mechanisms of speech perception. Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo 28, 67-73.

Fujisaki, H. and T. Kawashima. (1970) Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo 29, 207-214.

Glanzman, D. L. and D. B. Pisoni. (1973) Decision processes in speech discrimination as revealed by confidence ratings. Paper presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., April.

Liberman, A. M., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Liberman, A. M., P. C. Delattre, F. S. Cooper, and L. J. Gerstman. (1954) The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychol. Monogr. 68.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith. (1957) The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 54, 358-368.

Liberman, A. M., K. S. Harris, J. Kinney, and H. Lane. (1961) The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. J. Exp. Psychol. 61, 379-388.

Liberman, A. M., I. G. Mattingly, and M. T. Turvey. (1972) Language codes and memory codes. In Coding Processes in Human Memory, ed. by A. W. Melton and E. Martin. (New York: V. H. Winston).

Massaro, D. W. (1972a) Preperceptual images, processing time, and perceptual units in auditory perception. Psychol.Rev. 79, 124-145.

Massaro, D. W. (1972b) Preperceptual and synthesized auditory storage. Studies in Human Information Processing, Wisconsin Mathematical Psychology Program, University of Wisconsin, Madison 72-1.

Mattingly, I. G., A. M. Liberman, A. K. Syrdal, and T. Halwes. (1971) Discrimination in speech and nonspeech modes. Cog. Psychol. 2, 131-157.

McNeill, D. and B. Repp. (1973) Internal processes in speech perception. J. Acoust. Soc. Amer. 53, 1320-1326.

Pisoni, D. B. (1971) On the nature of categorical perception of speech sounds. Unpublished Ph.D. thesis, University of Michigan. (Issued as Supplement to Haskins Laboratories Status Report on Speech Research.)

Pisoni, D. B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. Psychophys. 13, 253-260.

Pisoni, D. B. and J. H. Lazarus. (1973) Categorical and noncategorical modes of speech perception along the voicing continuum. Unpublished manuscript.

Pollack, I. and D. B. Pisoni. (1971) On the comparison between identification and discrimination tests in speech perception. Psychon. Sci. 24, 299-300.

Sachs, R. M. (1969) Vowel identification and discrimination in isolation vs. word context. Quarterly Progress Report, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass. No. 93, 220-229.

Stevens, K. N. (1968) On the relations between speech movements and speech perception. Zeitschrift fur Phonetik, Sprachwissenschaft, und Kommunikationsforschung 21, 102-106.

Stevens, K. N., A. M. Liberman, M. Studdert-Kennedy, and S. E. G. Ohman. (1969) Cross-language study of vowel perception. Lang. Speech 12, 1-23.

Studdert-Kennedy, M. (1973) The perception of speech. In Current Trends in Linguistics, Vol, XII, ed. by T. A. Sebeok. (The Hague: Mouton).

Effects of Amplitude Variation on an Auditory Rivalry Task: Implications Concerning the Mechanism of Perceptual Asymmetries*

Susan Brady-Wood[+] and Donald Shankweiler[+]
Haskins Laboratories, New Haven, Conn.

Right-ear superiority in perception of dichotically presented words was interpreted by Kimura (1961), who discovered the phenomenon, and most subsequent workers as a manifestation of the specialization of the left cerebral hemisphere for language. This interpretation is supported by the finding of reversed ear asymmetry--left-ear superiority--in persons known on other grounds to have atypical, right-hemisphere speech representation. Work at Haskins Laboratories has shown that the right-ear advantage may be obtained with nonsense syllables that contrast in only one consonant segment. Since a right-ear advantage is not obtained from just any noise made by the human vocal tract, nor from acoustic fragments of speech sounds isolated from the syllable, it was concluded that even separate speech sounds are perceived by the dominant hemisphere because of their linguistic functions.

Right-ear superiority, as we understand it, requires that two conditions be met: 1) the stimulus material must require at some stage left-hemisphere processing (in general it has been found that nonspeech sounds do not); 2) the left ear's signal must undergo degradation in neural transmission to the left cerebral cortex which makes it less likely to be processed than the signal arriving at the right ear. Although direct evidence is lacking that the crossed auditory pathways are physiologically stronger in man, electrophysiological work on cat shows that each ear commands more neural units in the opposite cerebral hemisphere. Thus we hypothesize with Kimura that the left-ear signal's disadvantage is related to the fact that the crossed connections from ear to brain prevail over the uncrossed; thus the right ear's connection with the speech-dominant left hemisphere is a privileged one. Right-ear superiority, then, is attributed jointly to an advantage in transmission of signals conveyed directly by the crossed auditory pathway and to lateral specialization of portions of the left cerebral cortex for some aspect(s) of the speech process.

---

*Presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., April 1973, under the title: Effects of Attenuation of One of Two Channels on Perception of Opposing Pairs of Nonsense Syllables when Monotically and Dichotically Presented.

[+]Also University of Connecticut, Storrs.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

If the right-ear advantage can in part be attributed to the quantitative superiority of the crossed transmission line, then the ear advantage should be sensitive to manipulation in a regular and consistent way by varying the intensity of the two input signals. If this can be demonstrated, it may permit us to isolate, in a future experiment, the lateralized cortical, process-related component of the ear advantage from the transmission component reflecting variations in the efficiency of the crossed pathway. These two components conceivably vary independently, and they are confounded in usual measures of the ear advantage. Partialling out the effect due to transmission would clarify the interpretation of differences in the magnitude of the ear advantage for different phonetic classes and among different individuals for a given stimulus class.

METHOD AND PROCEDURE

In this experiment, we varied the intensity of the signal on one channel while holding the other signal at a constant level. Figure 1 shows the two conditions of competing stimulation which we compare here: the monaural or monotic, in which two synchronous signals are electrically mixed and then presented to one ear, and the dichotic condition, in which a different signal is presented to each ear.

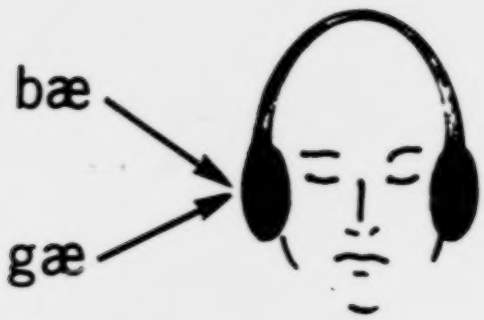
The syllables were the six stop consonants with the vowel /æ/. The stimuli were prepared on the Haskins Parallel-Resonance Formant Synthesizer. They were digitized, edited for amplitude level by a computer-assisted routine, and output into a test order containing synchronous pairs of random combinations of syllables. The overall amplitude was attenuated in 5-db steps from a reference level, which was chosen to be a comfortable listening level of about 70 db SPL. On each trial pair, the subject heard one syllable at the reference level and the other attenuated by 5, 10, 15, 20, or 25 db. The subjects, who were undergraduate psychology students, were informed of the stimulus set and given practice in identifying the syllables. They were asked to write down two different consonants for each trial, listing them in order of confidence. The data we present are based on the first response judgments.

RESULTS AND DISCUSSION

The data from the monotic experiment provide a baseline for evaluation of the dichotic condition. In Figure 2, the zero point represents percent correct identifications with no attenuation of either channel. The points to the left of the zero point indicate the percent correct identification of the stimuli on the attenuated channel; the points to the right of zero indicate the percent correct on the unattenuated channel for varying degrees of attenuation of the second channel. Each point is based on 30 judgments for each of 12 subjects. Ten db of difference in gain produced a nearly asymptotic change in identification performance. As expected, it is a matter of indifference whether we present these competing signals to the right ear or to the left ear.

Twenty-one subjects participated in the dichotic experiment in which twice as many (60) judgments per data point were collected. Figure 3 gives the plot of the data averaged for the 17 subjects who showed right-ear superiority at equal amplitude. Here the axes are the same: the plot shows the percent identification of stimuli presented to each ear at each level of attenuation relative to the opposite ear's signal. The effect of amplitude variation on performance is much less steep than was observed in the monotic condition. Thus, when

DICHOTIC



MONOTIC

Figure 1

Figure 2

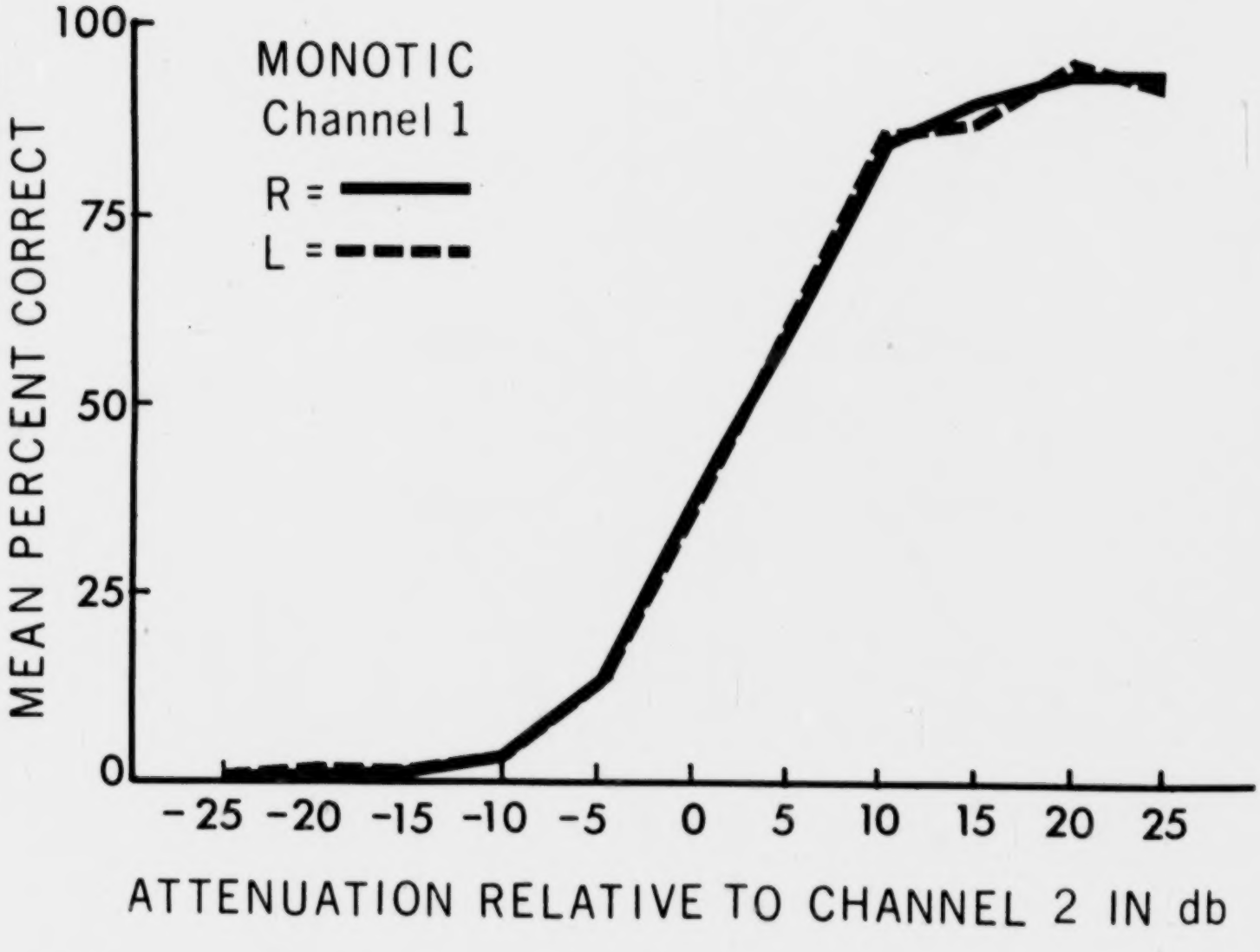
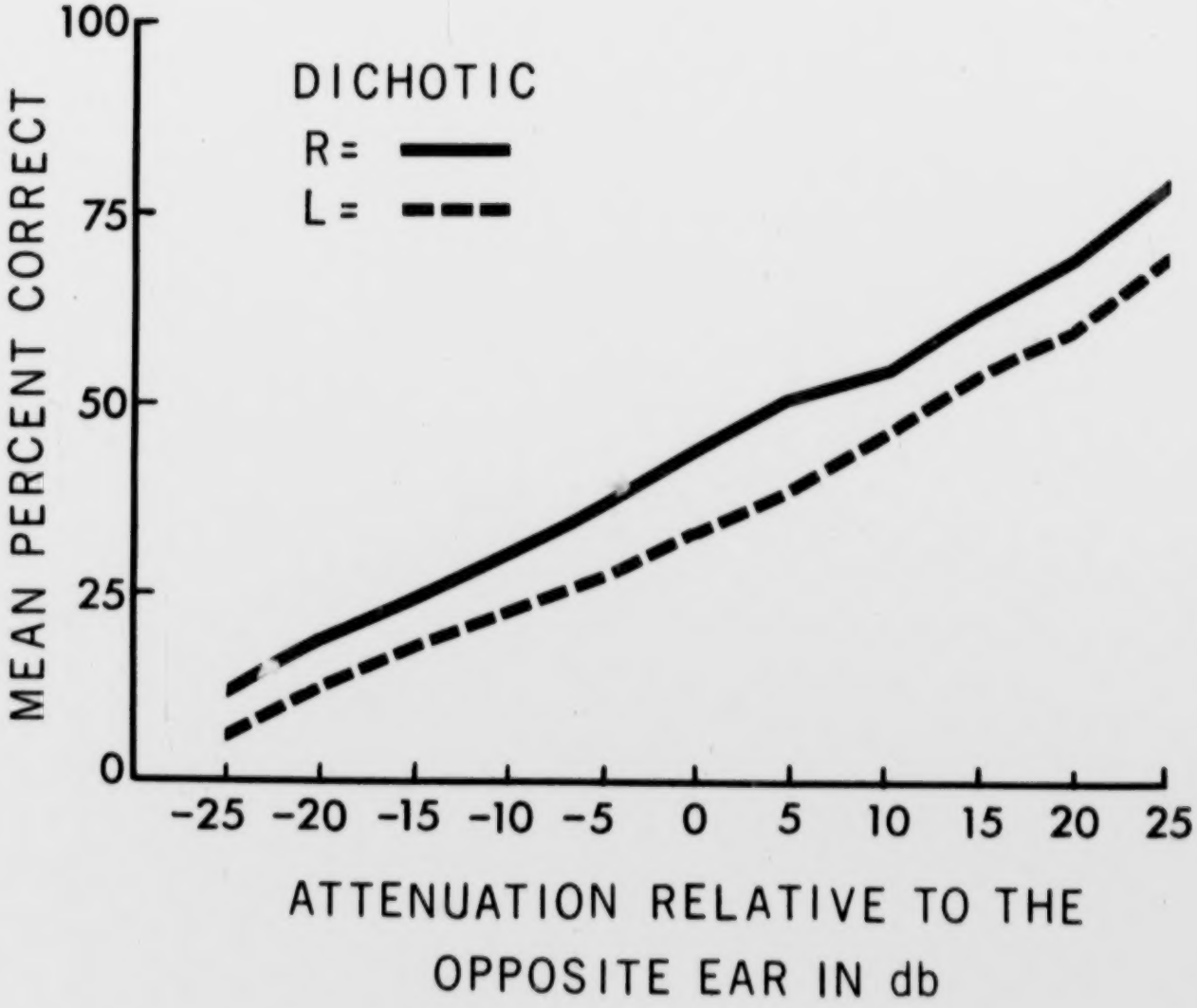

MONOTIC
Channel 1
R =
L =
100
75
50
25
0
MEAN PERCENT CORRECT
-25 -20 -15 -10 -5 0 5 10 15 20 25
ATTENUATION RELATIVE TO CHANNEL 2 IN db

Figure 3


DICHOTIC
R =
L =
MEAN PERCENT CORRECT
100
75
50
25
0
-25 -20 -15 -10 -5 0 5 10 15 20 25
ATTENUATION RELATIVE TO THE
OPPOSITE EAR IN db

the locus of competition is central, a given difference in amplitude biases perception far less than when the signals interact peripherally. The maximum amplitude difference of 25 db does not produce an asymptotic decrement in performance. Results obtained by Stafford (1971) at the Kresge Laboratory (New Orleans) place asymptote at between 40 and 50 db. It will be noted that the functions are parallel. The right-ear advantage remains at a constant 4 db relative to the left ear when it is attenuated by the same amount.

The 4 db difference appears as the cross-over point in Figure 4, which is simply the same data replotted to show the mean percent by ear for all trials of a given type. For example, the pair of points on the extreme right represents performance for those trials on which the left ear was attenuated by 25 db relative to the right, and the corresponding points on the extreme left gives performance for the reverse situation. The cross-over point, about 4 db from zero, shows the ear advantage which was displayed in Figure 3 as the difference between the parallel lines. This point varied for individual subjects from 1 to 14 db. There is a significant correlation of .80 between cross-over point and the degree of right-ear advantage when the signals are matched for amplitude. Variations in cross-over point reflect variations in left-ear gain required just to cancel the right-ear advantage. From this we may infer that individual differences in the magnitude of the right-ear advantage reflect, in part, differences in relative efficiency of the two transmission routes to the speech processor.

A further purpose of this experiment was to compare the effects of amplitude differences on perception of double stimuli with the effects of varying the relative time of onset. Drawing on earlier data for this comparison, we find similar effects of amplitude and time differences in the monotic case, different effects in the dichotic case.

A study by Studdert-Kennedy, Shankweiler, and Schulman (1970) introduced temporal onset asynchronies of 10 to 120 msec. With monotic presentation of stop-vowel syllables, temporally staggered by these amounts, a function very similar to that shown in Figure 2 was obtained. As is characteristic of peripheral masking, the advantage goes to the leading stimulus and the gain in performance as lead time increases is steep. In the dichotic case, the results were very different, but they are not parallel to those obtained when amplitude is varied dichotically. Interaural amplitude changes, as is seen in Figure 3, produce a linear effect on identification. By contrast, the effects of interaural differences in time of arrival are highly nonlinear. The lagging syllable has the edge in competition with the leading one; i.e., the masking effect is backward. Thus, in contrast to the findings shown in Figure 3, the plot giving identification as a function of onset asynchrony is asymmetric; performance changes more rapidly with lag than with lead. We may infer that interaural time differences affect the inputs after they have converged at the terminal cortical processor; interaural amplitude differences, on the other hand, affect the signals prior to entry to the cortical processor.

In concluding, we note that similar effects have been reported in studies investigating the parameters of visual masking of letters and words. Turvey (1973) found in studies employing a patterned mask that intensity and time manipulations produce different effects. In monoptic masking of a letter target by a pattern mask the relative intensities of target and mask are critical. By contrast, in the dichoptic situation--where interaction of the signals occurs only

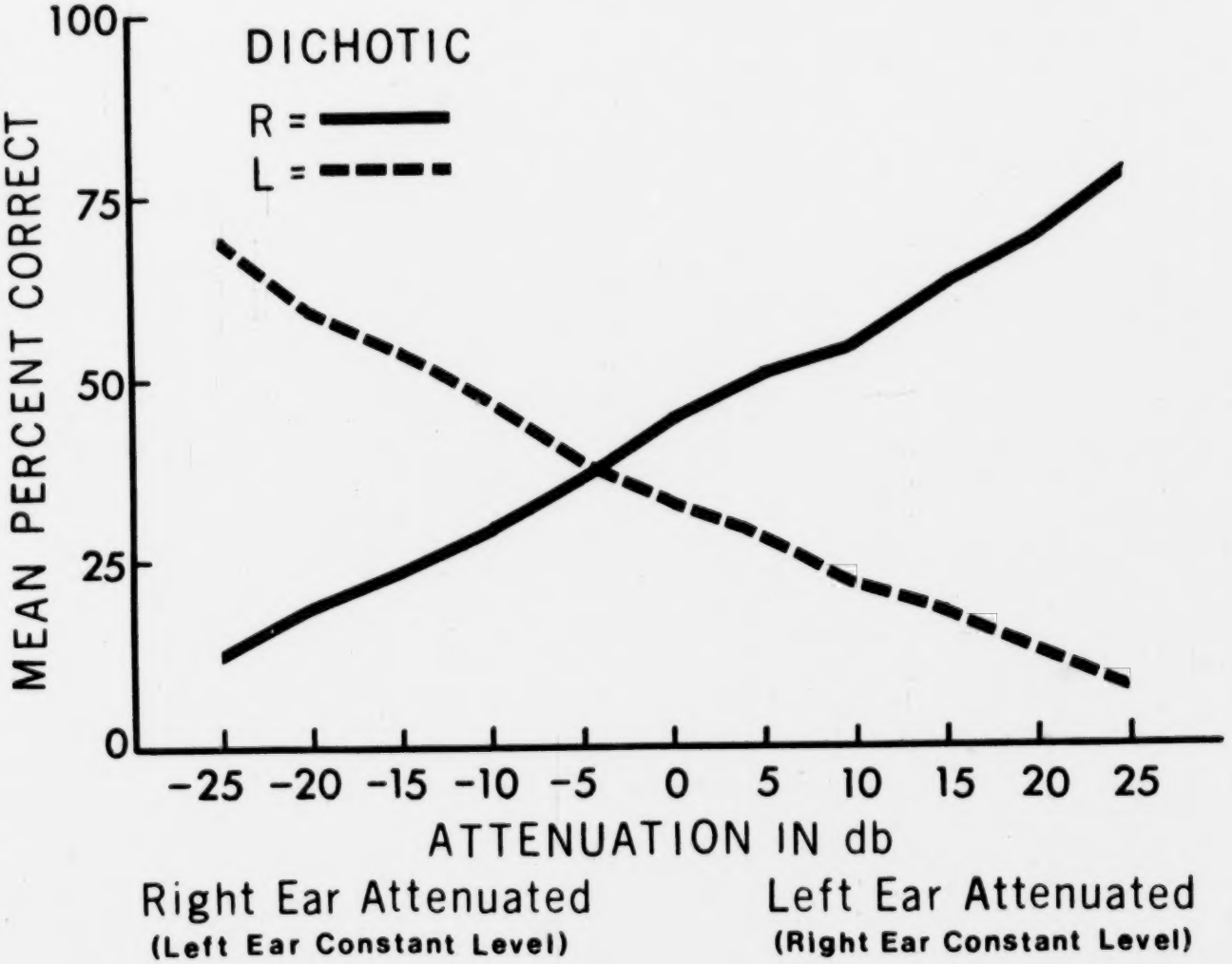

DICHOTIC
R =
L =
100
75
50
25
0
MEAN PERCENT CORRECT
-25 -20 -15 -10 -5 0 5 10 15 20 25
ATTENUATION IN db
Right Ear Attenuated
(Left Ear Constant Level)
Left Ear Attenuated
(Right Ear Constant Level)

Figure 4

at a central level--intensity differences are of little importance. We have shown a parallel difference in the effects of signal amplitude in monotic and dichotic listening, as is seen from a comparison of Figures 2 and 3. Finally, for both the auditory and visual modalities, the temporal direction of the central component of masking is asymmetric, being chiefly in the backward direction: processing of the first item is interrupted by a more recent input.

To summarize the principal findings of the present study, it was shown that when one signal is presented at a fixed amplitude and the competing signal is varied in 5 db steps, the function relating identification to the degree of attenuation is linear (for amplitude differences at least as great as 25 db) and relatively flat in the dichotic case. From the effects of attenuation on the ear advantage, it was found that variations among individuals in the magnitude of the right-ear superiority are in part determined by factors related to transmission of the auditory signal prior to cortical processing. This, we hope, will make possible the development of ways to isolate the two components of the ear advantage.

REFERENCES

Kimura, D. (1961) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Stafford, M. (1971) Dichotic speech perception with interaural intensity differences. Unpublished Masters thesis, Tulane University, New Orleans, La.

Studdert-Kennedy, M., D. Shankweiler, and S. Schulman. (1970) Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. J. Acoust. Soc. Amer. 48, 599-602.

Turvey, M. T. (1973) On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. Psychol. Rev. 80, 1-52.

Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects*

Ruth S. Day+
Haskins Laboratories, New Haven, Conn.

Ordinarily individuals fall into a normal distribution in perception and memory experiments. That is, when we plot the number of subjects who achieve high, medium, and low scores in a task, we find that most fall in the middle range, while some fall at the high and low ends. Such a distribution occurs for a wide variety of scores, including for example, percent correct identification and number of trials to criterion.

DICHOTIC FUSION STUDIES

Phonological fusion tasks in dichotic listening consistently do not yield a normal distribution (Day, 1969). A series of tasks is given to the same subjects, using the same dichotic tapes in each task. All items are of the general form BANKET/LANKET, where the inputs to each ear differ only in their initial phoneme. Furthermore, these initial elements can be fused into a cluster in English. Thus BANKET/LANKET⟶BLANKET.[1]

Identification task. Subjects are asked to report 'what they hear' on every trial, be it "one word or two...a real word or a nonsense word." Often subjects report hearing BLANKET, which is a fusion response: the /b/ and /l/ are sent to different ears yet are perceived as a fused cluster. Figure 1 shows the frequency distribution of fusion scores for a typical group of subjects; fusion scores are expressed as the proportion of trials on which each subject fused. Clearly, the distribution is bimodal, with high fusers and low fusers at either end and a marked absence of subjects in the middle range. Although Figure 1 shows data for only 16 subjects, the bimodality has occurred over hundreds of subjects in subsequent experiments.

Temporal order judgment (by phoneme). One could argue that the identification task emphasizes processing at the word level. However, elsewhere (Day, 1968) it has been shown that fusions also occur when both inputs are acceptable English words (e.g., BACK/LACK⟶BLACK), as well as when the fusion is a nonword

---

*Paper presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., 11 April 1973.

+Also Yale University, New Haven, Conn.

[1]The arrow should be read "yields."

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

Figure 1

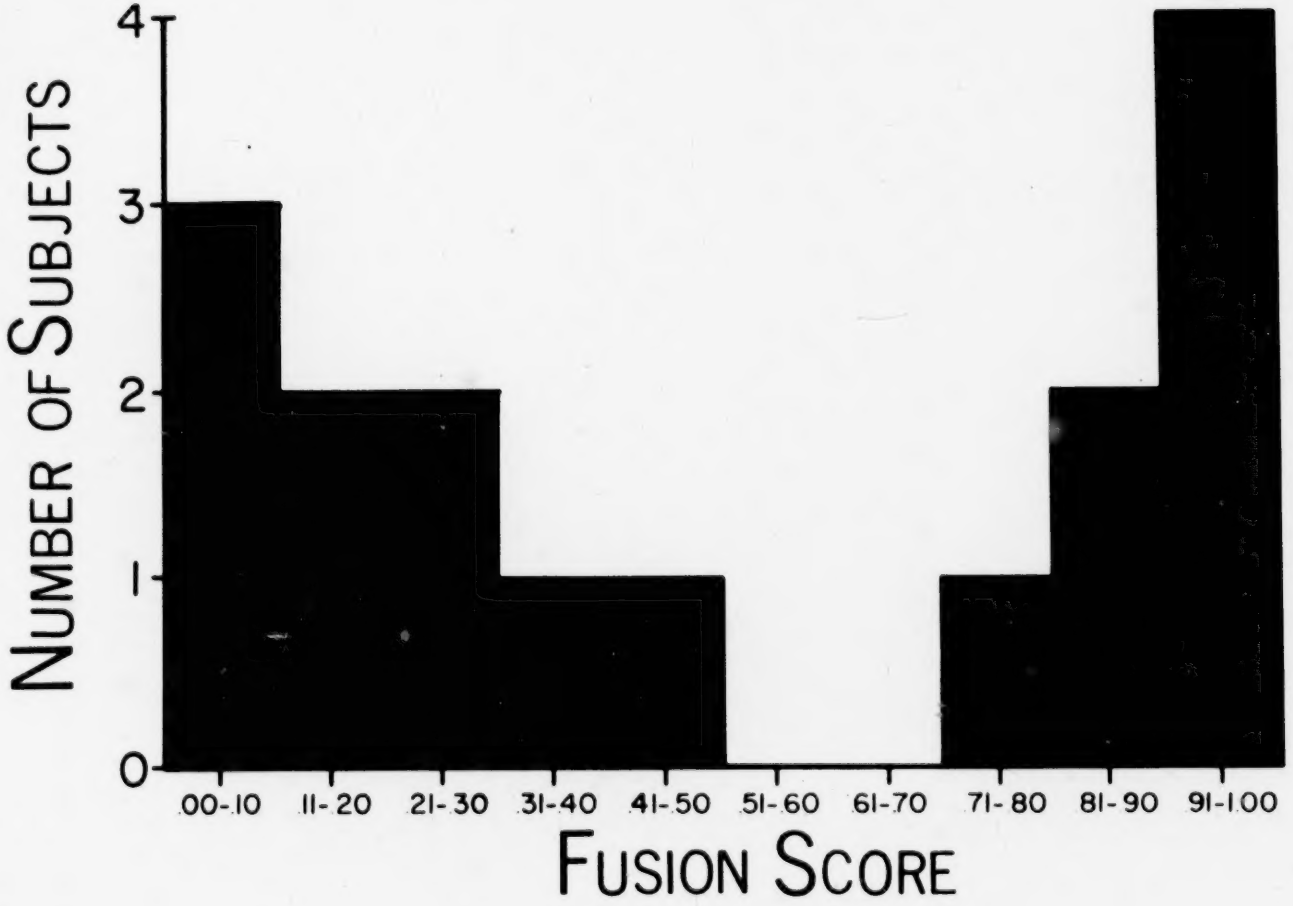


Figure 1: Frequency distribution of fusion rates over subjects. (From Day, 1969)

(e.g., GORIGIN/LORIGIN⟶GLORIGIN). Nevertheless, another test is needed, a test which de-emphasizes combining all the information into a single response. A temporal order judgment task meets this requirement (Day, 1969). The same subjects listen to the same fusible dichotic items, but this time they need to report only the leading phoneme in each pair. On half the trials, the stop consonant (e.g., /b/ in BANKET) begins first by a short interval while on the remaining trials the liquid (e.g., /l/ in LANKET) begins first. Subjects are asked to report 'the first sound (phoneme) they hear,' that is, to make a temporal order judgment (TOJ).

Typical TOJ-by-phoneme results are shown in Figure 2. At the top of the display is a subject who was correct in determining temporal order when the stop consonant led, but incorrect when the liquid led. Note that in English, stop + liquid can occur in initial position, but liquid + stop cannot occur initially. Hence, this subject reported what the language allows, not what the leading phoneme was. Another type of subject is shown at the bottom of Figure 2. This subject was highly accurate in judging the temporal order of fusible items, no matter whether the stop or the liquid led. In summary, the first subject is a poor judge of temporal order while the second is a good judge. Note that the TOJ-by-phoneme task requires only phonetic processing of the initial stop and liquid; it does not require phonological processing of these units into a cluster. Nevertheless, some subjects seem unable to disengage phonological processing mechanisms.

Relationship between fusion and TOJ-by-phoneme. So far we have considered two tasks and have found contrastive performance among individuals in each. In the identification task, there were high and low fusers, while in the TOJ-by-phoneme task there were good and poor judges of temporal order. The correlation between performance on the two tasks is shown in Figure 3. The high fusers are poor judges of temporal order and have been termed "language-bound," since they are heavily influenced in both tasks by the phonological rules of the language. The low fusers are good judges of temporal order and have been termed "stimulus-bound," since they are highly accurate in reporting facts about the stimulus conditions.

Temporal order judgment (by ear). A final task using the same subjects and tapes again requires a TOJ. However, this time subjects are asked to report "which ear led" on each trial. Such a judgment does not require linguistic processing since the subject need not identify any phonemes. There is dramatic improvement in this task: some language-bound subjects show increases in performance as great as 50% or more on liquid-leading trials. Nonetheless, some still perform better when the stop consonant leads, suggesting that they have not totally disengaged phonological processing mechanisms.

Discussion. The general strategy in these dichotic fusion studies is to determine the level at which an individual can disengage linguistic processing mechanisms. Some do it readily, while others have great difficulty in doing it.

The individual differences in the dichotic fusion tasks appear to be stable over time. Language-bound subjects given extended practice on the TOJ-by-phoneme task may improve a small amount, but still show grossly inferior performance in comparison with stimulus-bound subjects (Day and Thompson, in preparation). These qualitative differences in perception appear to depend upon the extent to

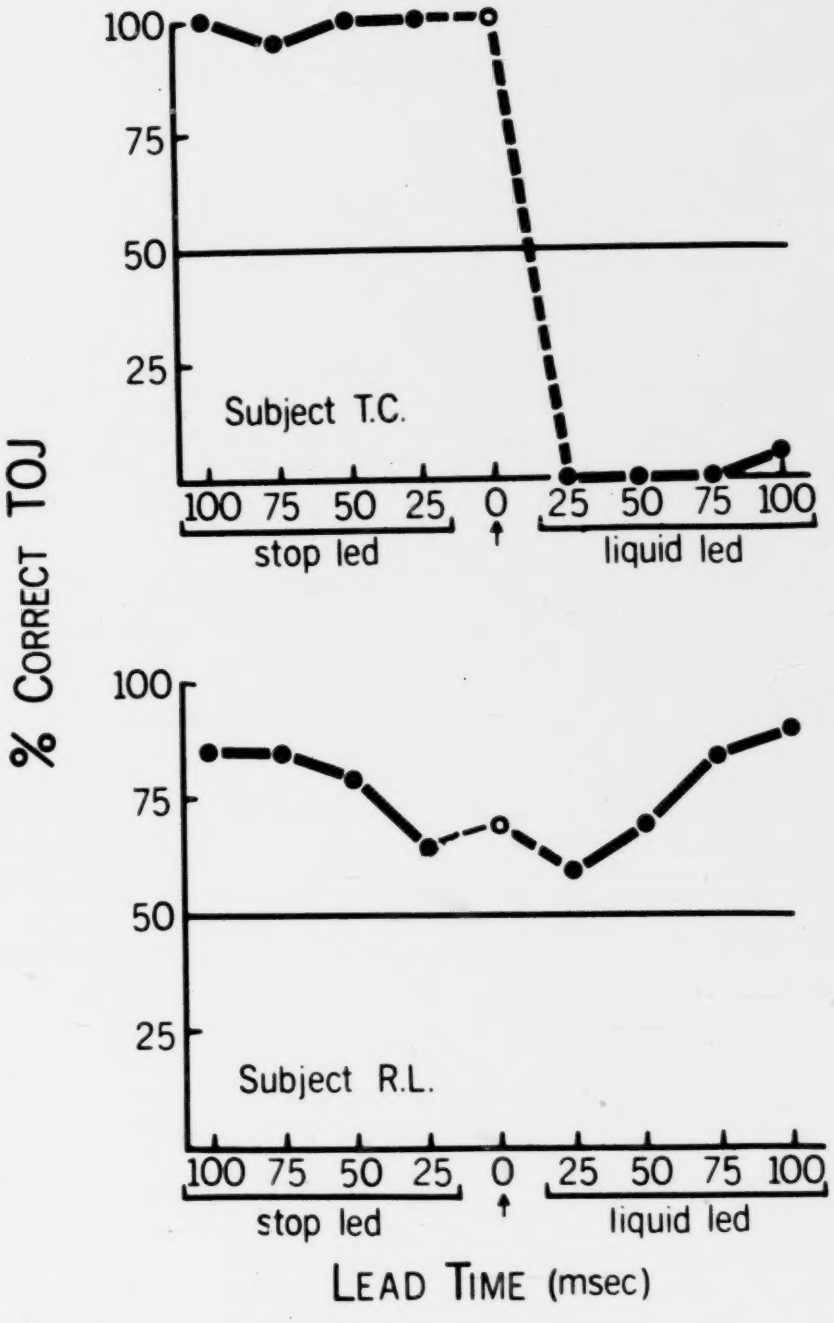


Figure 2: Percent correct temporal order judgment (TOJ) for dichotic trials where the leading phoneme was either a stop consonant (e.g., /b/) or a liquid (e.g., /l/). Each display was obtained from a single subject. (From Day, 1969)

Figure 3

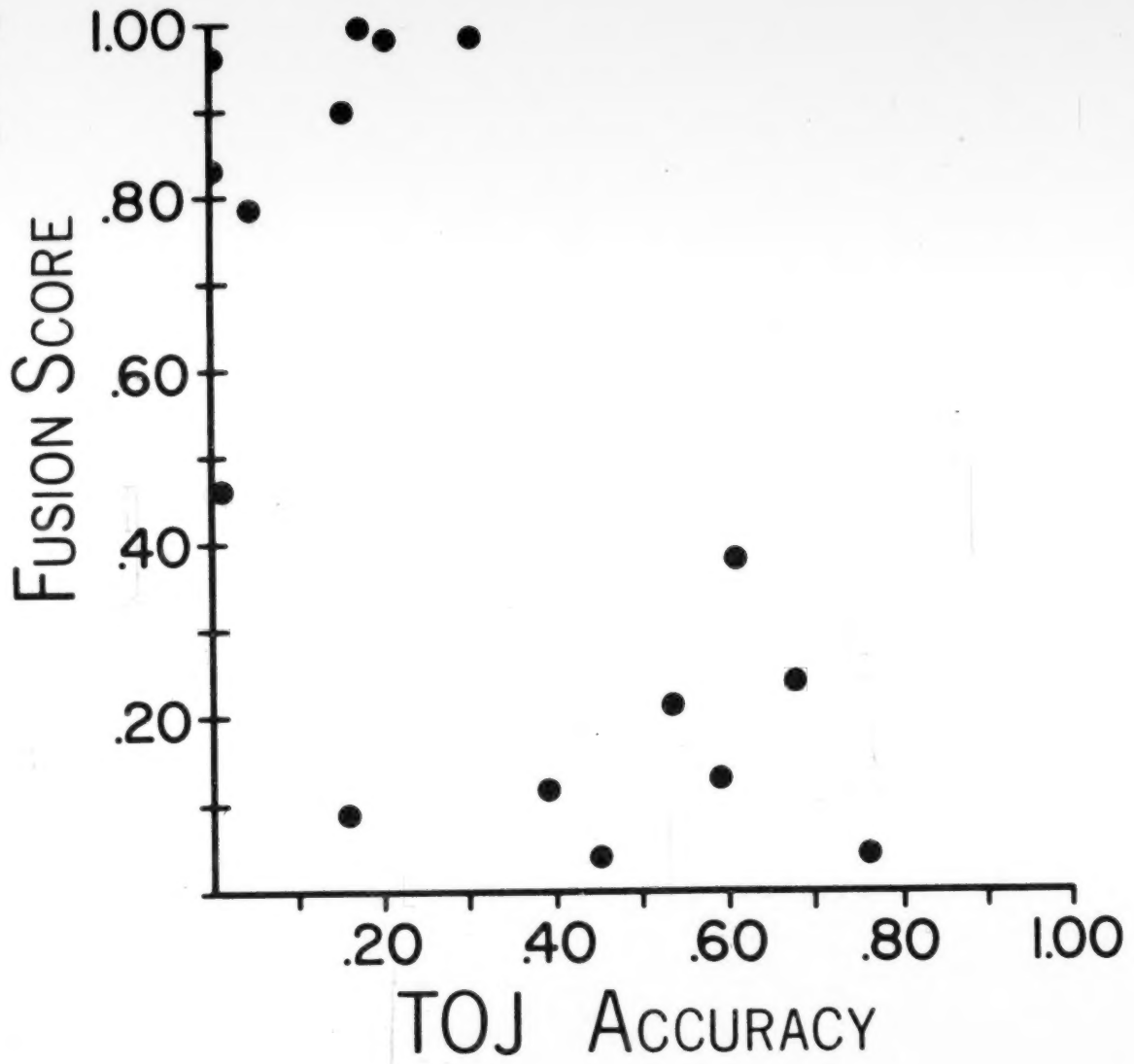


Figure 3: Relationship between fusion rate and temporal order judgment accuracy for the same subjects. (From Day, 1969)

which subjects are influenced by various linguistic constraints, even when reference to these constraints is not necessary for task solution.

It is interesting that such dramatic differences occur in the dichotic fusion tasks. However, if the differences occur only within these tasks, then they are of limited interest. The major question, then, is: do subjects retain their group identity--language-bound, stimulus-bound--on other cognitive tasks? The present paper explores this question using the digit-span memory task.

DIGIT-SPAN MEMORY

In a typical digit-span test a series of numbers is presented in rapid succession, for example, 3-2-6-8-5-4-7-1-9. The subject's task is to report all the numbers in the exact order they were presented. He is given an answer sheet with a horizontal array of blanks; from left to right, these blanks represent the temporal order of the items. In order for a digit to be scored correct it must be placed in the appropriate blank. Some typical results are shown in Figure 4, which plots percent correct for each serial position, from the first item to the last in the list. Performance is best at the beginning and end of the list, and falls in the middle. These data represent the "classical serial position curve." Superior performance at the end of the list is called the "recency" effect; items here appear to be in a very temporary storage system. Superior performance at the beginning of the list is called the "primacy" effect; items here appear to be in a more permanent storage system. Such results have been reported many times in the psychological literature.

Control condition. In the present experiment, language-bound and stimulus-bound subjects were selected on the basis of their performance on the dichotic fusion tests. The subjects were then given a series of digit-span lists; each list contained nine digits and was spoken at a rate of two digits per second. The data obtained are those already shown in Figure 4. These data, however, have been averaged over the two groups of subjects. Figure 5 shows the same data plotted separately for each group, and shows clear, quantitative differences between the two groups: overall performance collapsed across all serial positions was 88% correct for stimulus-bound subjects and 63% correct for language-bound subjects (Mann-Whitney $U = 8.5$, $p < .001$).

Stimulus-bound subjects showed little evidence of a serial position effect, while language-bound subjects showed a very marked serial position effect (with performance dropping as low as 28% correct for the middle item). Despite these apparent differences in the shapes of the two curves, caution must be exercised in concluding that there are qualitative differences in memory. The large difference in overall performance level between the two groups makes such inferences difficult from a statistical point of view.[2]

It is important to emphasize the contrast between the averaged data (Figure 4) and the data separated by groups (Figure 5). If indeed language-

[2] It could be that both groups display the serial position effect, but that high performance "ceiling effects" are concealing it in the stimulus-bound data. This problem will be reconsidered later.

Figure 4

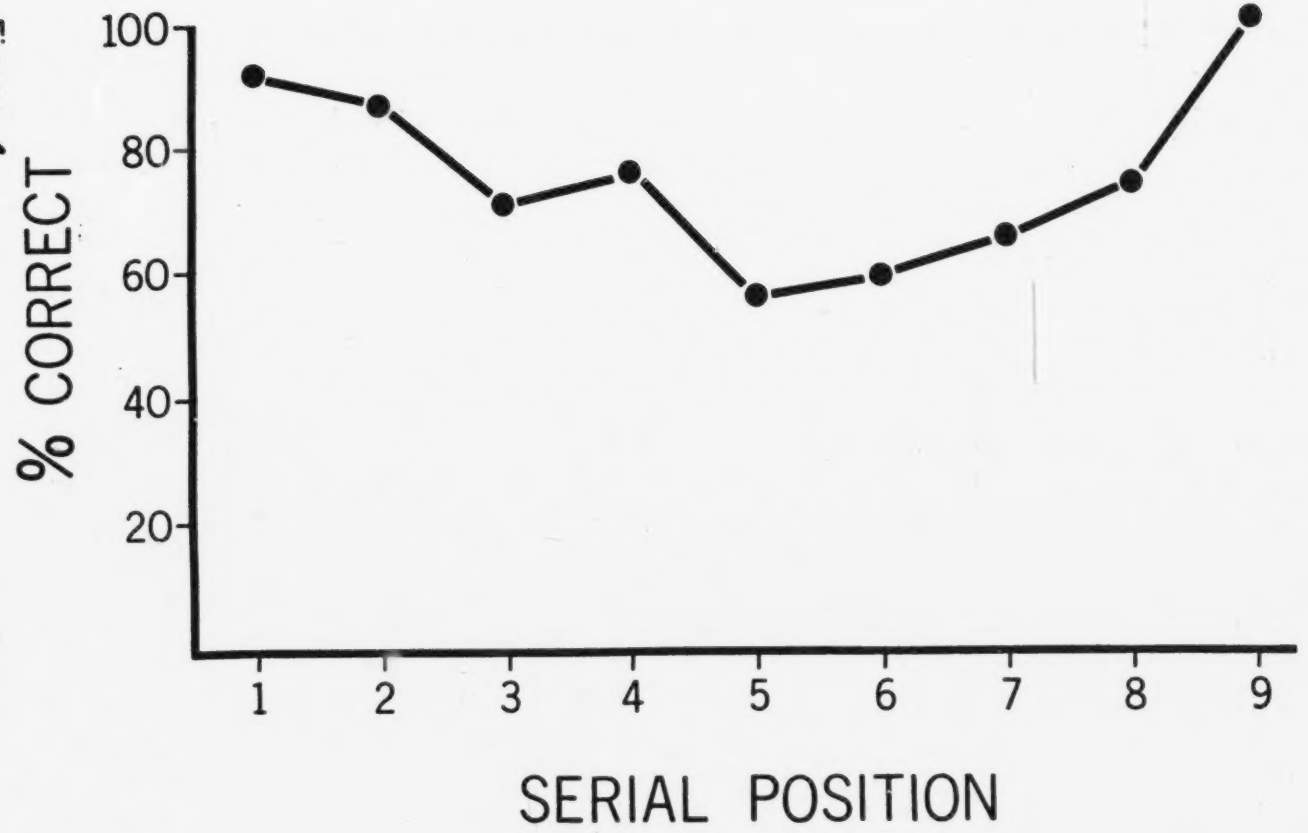


Figure 4: The classical "serial position effect" in digit-span memory experiments.

Figure 5

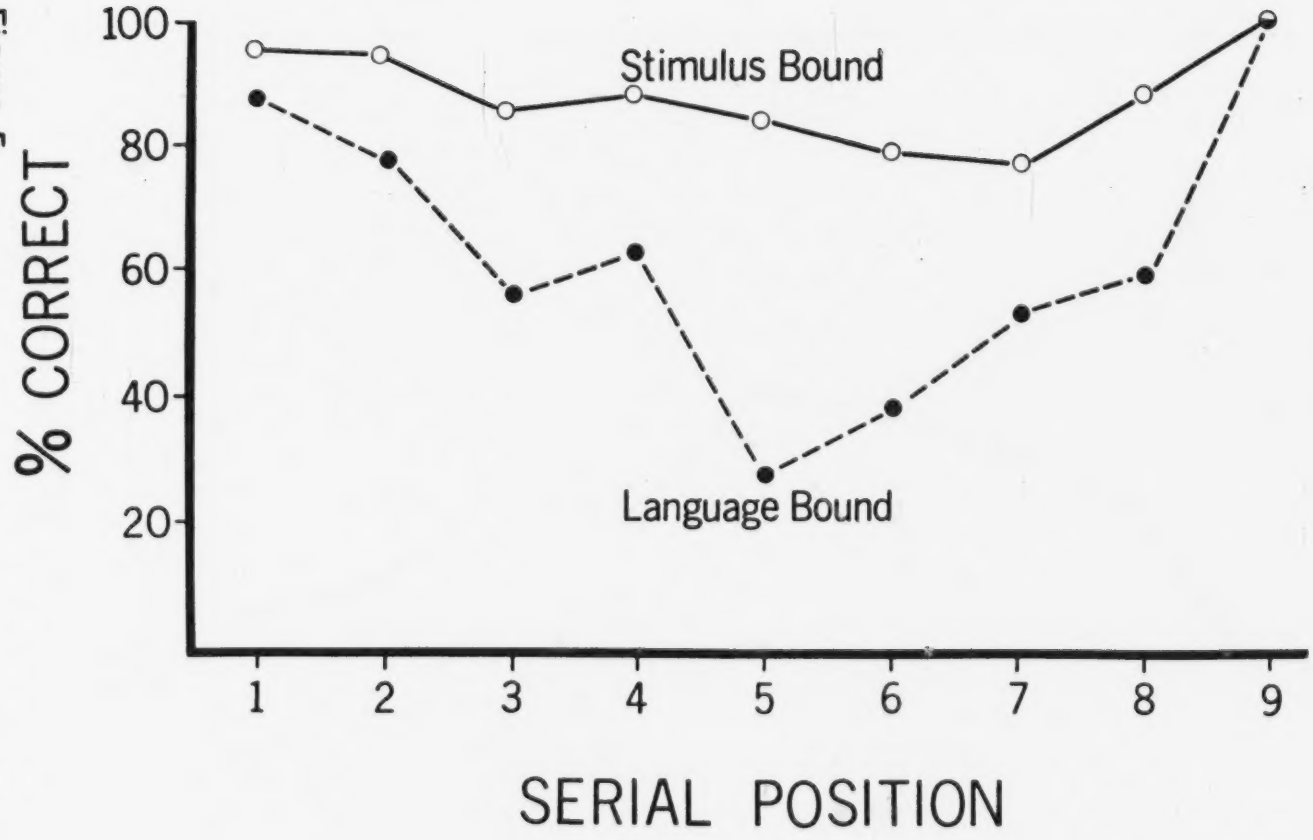


Figure 5: Digit-span memory for language-bound and stimulus-bound subjects (replotted from Figure 4).

bound and stimulus-bound characteristics are representative of the population at large, then the classical serial position curve could be one that all experimenters obtain but that no individual subjects display. Hence we could be building models for performance that does not occur.

Suffix condition. A variant of the digit-span task was also given. At the end of each list the word "zero" was added. Subjects were told that they did not need to report the zero, but to use it as a cue to begin their recall. Thus the zero acts as a redundant suffix. Crowder and Morton (1969) have shown that the suffix hurts items at the end of the list, the items that are in a very temporary storage system. Items elsewhere in the list are not affected by the suffix. In Crowder and Morton's terms, the suffix harms only those items in a "precategorical acoustic store."

Data from the suffix condition of the present experiment are shown in Figure 6, where they are contrasted with the data of Figure 5 from the control (no suffix) condition. For language-bound subjects, the suffix hurt performance only for items at the end of the list. For stimulus-bound subjects, however, the suffix reduced performance throughout the list; nevertheless their overall performance remained very high. The differential effects of the suffix on the two groups suggest that we may be dealing with qualitative differences in memory.

The suffix data again appeared to show differences in the shapes of the serial position curves for the two groups. The stimulus-bound curve was approximately flat, while the language-bound curve showed the usual serial position effect, although with the recency effect weakened by the suffix. However, again if is difficult to make shape comparisons since the overall level of performance was so different for the two groups.

Overall performance transformations. In order to make between-group comparisons of curve shape more meaningful, the data from both the control and suffix conditions were transformed to take into account the differences in overall performance level. For each group in each condition, percent correct for a given serial position was divided by the overall percent correct on that entire test. The results of these transformations are shown in Figure 7. If there were no differences in memory for the various portions of the list, then the resulting figures ought to be about 11% at each of the nine serial positions, regardless of overall performance level. Stimulus-bound subjects clearly showed this "flat" pattern of results in the control condition, and a pattern very close to it in the suffix condition.[3] Language-bound subjects, in contrast, continued to show serial position effects in both conditions. Since there are clear differences in curve shape between the groups in both conditions, even when differences in performance level are taken into account, the possibility that language-bound and stimulus-bound subjects possess qualitative differences in memory receives additional support.

---

[3] If ceiling effects were concealing the serial position effect in the stimulus-bound control condition, as suggested earlier, then the suffix data ought to show evidence of the classical curve, since there were sufficient overall errors to yield a differential distribution over the beginning, middle, and end of the list. Since both stimulus-bound curves are essentially flat, the ceiling effect argument is weakened. Instead, it appears that these subjects have comparable memory levels over all portions of the list.

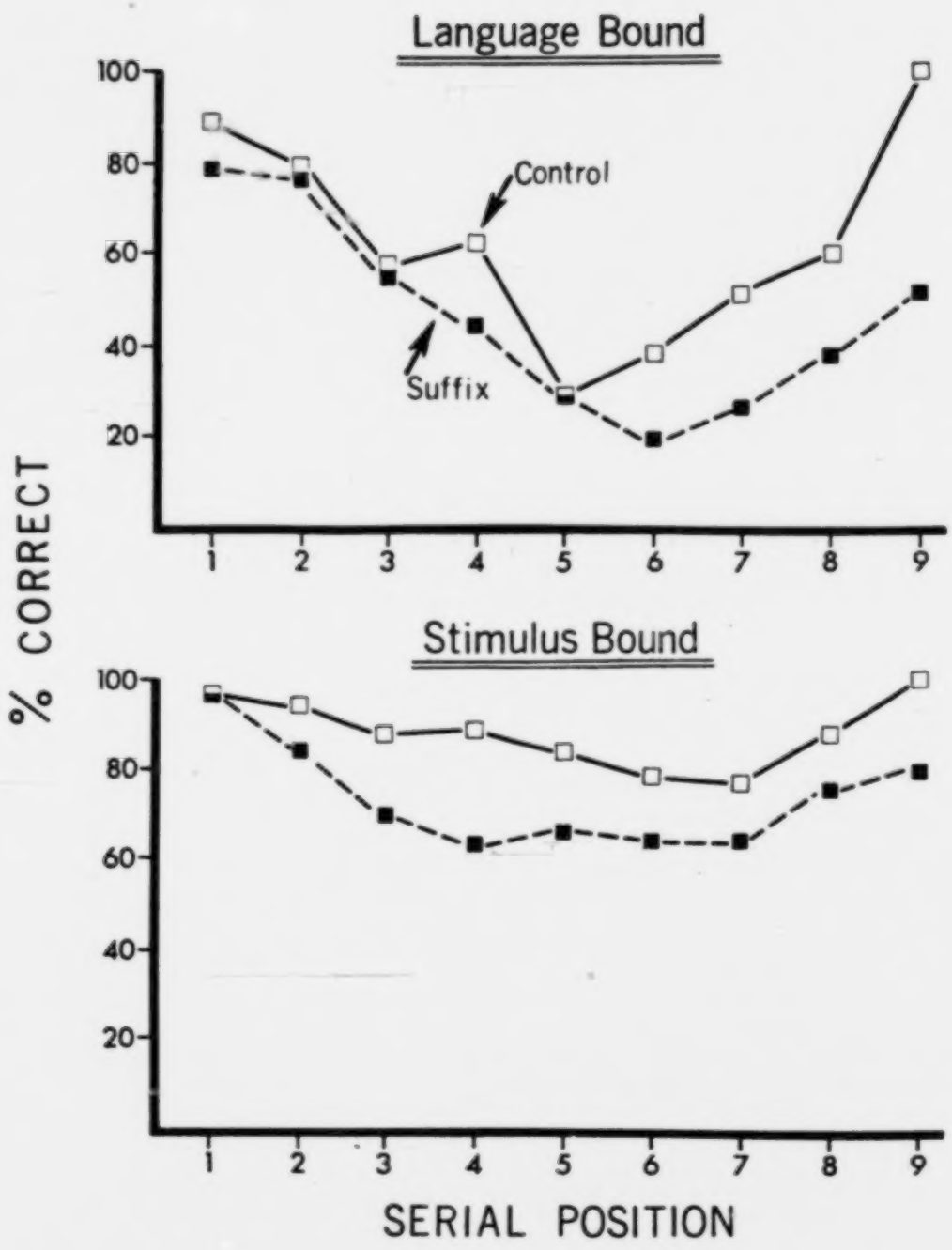


Figure 6: Control vs. suffix conditions for language-bound and stimulus-bound subjects.

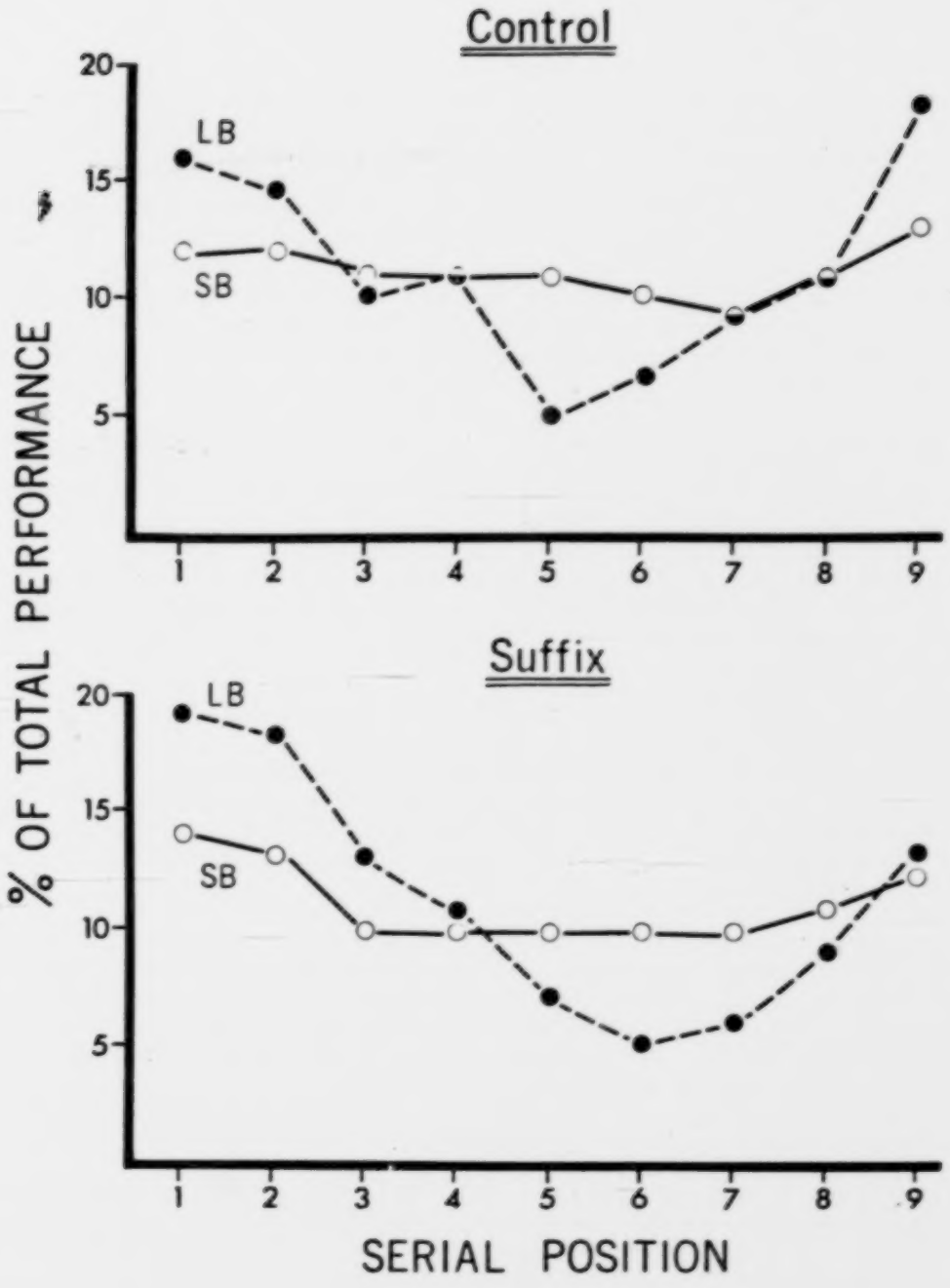


Figure 7: Data of Figure 6 transformed to control for overall performance levels.

DISCUSSION

What is responsible for the group differences in digit-span performance? A satisfactory explanation will have to wait until further parametric variations have been completed (for example, those dealing with rate of presentation and list length). However, some speculations can be made on the basis of the present data.

One code or two? Stimulus-bound subjects may translate the spoken digit string into a companion visual representation. The visual image of the string would be arrayed spatially. It is well known that spatial tasks are handled by the nonlanguage hemisphere of the brain, which for most people is the right hemisphere. Hence stimulus-bound subjects would be employing the spatial processing abilities of the right hemisphere in addition to the verbal processing capabilities of the left hemisphere. The existence of two storage codes, visual and verbal, for the digit strings could facilitate overall retention. While there is no direct evidence for this interpretation from the present experiment, elsewhere we have shown that stimulus-bound subjects are superior on a visual search test that emphasizes spatial flexibility (Day, in preparation).

Duration of storage systems. Another way to view the present data is to consider the duration of various storage systems. Most current models of immediate memory include an initial store in which spoken items are held in a relatively unanalyzed form, for example, the precategorical acoustic store (PAS) of Crowder and Morton (1969). Items must undergo phonetic processing in order to enter short-term memory (STM) where about seven (plus or minus two) chunks can be held for several seconds (Miller, 1956). Stimulus-bound subjects may be able to transfer items from PAS more quickly, and hence be able to "read out" all the items from a single store, STM, once stimulus presentation has been completed. Language-bound subjects may have a slower transfer system, so that when stimulus presentation is completed, some (early) items are in STM, other (late) items are still in PAS, and the remaining (middle) items are in transition between the two systems. Read-out of items from STM is no problem, and items in PAS are followed through to STM and reported accurately. However, the "transition" items often get lost. Experiments currently under way on rate of digit presentation will help clarify whether there are differences in processing speed between language-bound and stimulus-bound subjects.

Conclusion. Language-bound and stimulus-bound subjects showed clear differences in a short-term memory task. These differences are at least quantitative, and may also be qualitative. They suggest that the two groups of subjects identified in the dichotic fusion tasks are not simply byproducts of a particular experimental situation. Instead they may represent differences that play an important role in general cognitive functioning.

REFERENCES

Crowder, Robert G. and John Morton. (1969) Precategorical acoustic storage (PAS). Percept. Psychophys. 5, No. 6, 365-373.

Day, Ruth S. (1968) Fusion in dichotic listening. Unpublished Ph.D. thesis, Stanford University (Psychology).

Day, Ruth S. (1969) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Paper presented at the 9th meeting of the Psychonomic Society, St. Louis, Mo., November. [Also Haskins Laboratories Status Report on Speech Research (1970) SR-21/22, 71-87.]

Day, Ruth S. (in preparation) Differences between language-bound and stimulus-bound subjects in solving word-search puzzles.

Day, Ruth S. and Elaine E. Thompson. (in preparation) Long-term practice with fusible dichotic items.

Miller, George A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev. 63, 81-97.

On Learning "Secret Languages"*

Ruth S. Day+
Haskins Laboratories, New Haven, Conn.

OO-DAY OO-YAY OH-NAY UHT-WAY IS-THAY EZ-SAY? For many people, the answer to this question is ES-YAY. DUH-GOO YUH-GOO NUH-GO WUH-GUT THUH-GIS SUH-GEZ? For most people, the answer to this question is I HAVEN'T THE FOGGIEST. Both utterances are examples of "secret languages": the first is very common and is known as Pig Latin, while the second is rare and is known as G-language.[1]

Secret languages are known to occur in many of the world's languages. Children often invent them to talk among themselves without comprehension by adults or by other children not in their immediate group. In the Philippines, courting adolescent couples have difficulty achieving physical intimacy, as they are closely watched by their chaperone; hence they use secret languages to gain verbal intimacy (Conklin, 1956). In Surinam, small groups of teenage boys or young men use secret languages to establish peer group solidarity (Price and Price, in press). In some cultures, skill in linguistic play is highly valued and is used more for entertainment's sake than for concealment, as in the "talking backwards" language of the Cuna Indians in Panama (Scherzer, 1970).

Secret languages usually begin with the native language and add a few new rules. In Pig Latin the basic rules are:

1. for each word, delete the first consonant (or consonant cluster)[2]
2. utter the remainder of the word
3. add the deleted consonant, followed by the vowel AY.

Therefore the word SECRET becomes EEKRUT-SAY. In G-language, the rules are:

---

*Paper presented at the Eastern Psychological Association meeting, Washington, D. C., 3 May 1973.

+Also Yale University, New Haven, Conn.

[1]The author's father is gratefully acknowledged for inventing G-language; it has yielded interesting data in the secret language experiment as well as some enjoyable family communication.

[2]There are additional rules for items that begin with vowels.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

1. for each syllable, utter the first consonant (or consonant cluster)[2] followed by the vowel UH
2. add G before the next vowel and continue with the rest of the syllable.

Therefore the word SECRET becomes SUH-GEE-KRUH-GUT.

Secret Language Experiments

Recently, we have been devising new secret languages that add rules at various levels of linguistic analysis. We then teach these languages to adults to see the extent to which they can operate on language at different linguistic levels. The following is a recorded passage in one of these new secret languages; see if you can determine what the rules are:

> HERRO. THIS IS A TLANSFOLMED RANGUAGE. HOPEFURRY, YOU WIRR BE ABER TO SPEAK IT. IT'S NOT VELY HALD TO DO, ONCE YOU FIGULE OUT HOW TO DO IT. UNTIR YOU KNOW THE LURES THOUGH, YOU WIRR HAVE TLOUBER UNDELSTANDING IT.[3]

This R-L language was devised for experimental purposes. Its rules are:

1. every time there is an /r/, change it to /l/
2. every time there is an /l/, change it to /r/.

Thus ROCKET becomes LOCKET, LAVISH becomes RAVISH, and CASSEROLE becomes CASSELORE. Note that R-L language involves new rules solely at the phoneme level; there are no changes at the syllable level as in most secret languages.[4]

In a typical word translation experiment, the subject is given a standard English word and asked to translate it into the secret language version. The following tape recorded passage taken from an experiment session illustrates the procedure:

---

[3] To facilitate reading, secret language utterances are given in orthographic rather than phonemic notation. Note that this notation does reflect the phonemic form of the secret language transformation rather than being a strict letter replacement system. For example, the phonemic representation of the standard and secret versions of the word TROUBLE are /trʌbəl/ and /tlʌbəɹ/; the notation for the secret language form is TLOUBER rather than TLOUBRE which would be pronounced /tlʌbrə/. When phonemic notation is needed, it is given between slashes.

[4] Phoneme substitutions do occur in some secret languages, for example le bolite spoken in Haiti (Alexis, 1966).

STIMULUS	SUBJECT C.F.
LOCKET	ROCKET
CASSEROLE	CASSELORE
PICKLE	PICKER
MIDDLE FINGER	MIDDER FINGLE
LIVER	RIVEL
LAWYER	ROYAL
NELSON ROCKEFELLER	NERSON LOCKEFERREL

This subject responded quickly and accurately; she had no difficulty making the appropriate substitutions. After the word transformation task was completed, she was asked to produce some connected discourse in secret language form. We asked for a well-known passage in order to decrease task demands; the passage was "Mary had a little lamb." Here is the same subject reciting this nursery rhyme in secret language form:

> MAILY HAD A RITTER RAM, ITS FREECE WAS WHITE AS SNOW. AND EVLIWEL THAT MAILY WENT, THE RAM WAS SHULE TO GO.

Note that all /r/'s and /l/'s were transformed appropriately, and that the whole passage was produced in a highly fluent fashion: pacing and intonation resembled those of ordinary speech.

Now consider another subject working on one item in the word translation task:

EXPERIMENTER	SUBJECT D.Q.
BRAMBLE	BRAIR...BRAIR...RORE
Not quite...BRAMBLE	...BERLAIRM...LULL...LORE
Not quite...	BERLER...BLERM...BULL...BULL
What are the two rules?... R goes to L...	...R to L, and L to R
...BRAMBLE	BLER...BLERM...BUR
That's close. BRAM would be?...	...BLERM
BLAM...You're just transforming R's and L's, okay?	Right. And BRAMBLE, I'm transforming the R to L. So it's BLER... BLER...BLERM. And BULL is...the L to the R...it's BER.
...You're still sticking an R in; you said BLERM. Where do you get BLERM? It was BRAM...	Yah, BRAM. And the R in BRAM goes to L. So it's BL--...BLERM.

BLAM! Just plain BLAM. See, you're still putting an R in next to the M...	BRAM, but you have B-R. And the R goes to L. So I'm trying to delete that R, and make it B-L rather than B-M.
...Rather than B-R.	Rather than B-R, right. So I'm trying to say BLER...BLERM.
Where do you get the R? It's BRAM originally, B-R-A-M...	B-R-A-M. Doesn't that go to B-L-A-M?
Right! How do you pronounce B-L-A-M?	BLAM.
Right. Okay, let's go ahead...	

This subject had great difficulty with the task. Even though he could state the two rules, he was unable to employ them effectively. He worked on this item for almost 3 minutes and never successfully transformed the whole word; finally, the experimenter went on to the next item. I will spare you his NELSON ROCKEFELLER.

These two subjects illustrate the wide range in ability that individuals have in learning the R-L language. The topic of individual differences will be discussed more fully later.

The Present Secret Language Experiment: General Findings

Method. In the present experiment, 63 Yale University students learned R-L language. After hearing a brief recorded passage in secret language form (as given above), they were told the two rules. They then took a word translation test. All 24 stimulus items were acceptable English words, but half yielded words (W) and half nonwords (NW) in their secret language versions. Sample items for the W⟶W[5] case were ROCKET⟶LOCKET and LAWYER⟶ROYAL. Sample items for the W⟶NW case were BRAZIL⟶BLAZIR and LIVER⟶RIVEL. There were 43 target phonemes (/r/ and /l/) in all, since some items contained multiple targets (e.g., BRAZIL). Subjects were instructed to transform only the /r/'s and /l/'s, and to keep all other sounds constant.[6] After completing the word translation task, subjects recited "Mary had a little lamb" in its R-L version. Each subject was tested individually. All responses were tape recorded and later transcribed into phonemic notation.

---

[5]The arrow should be read "yields."

[6]Admittedly, there are some obligatory phonetic changes in neighboring units as /r/ and /l/ replace each other. For example, the vowel before the final liquid must be qualitatively different in BRAZIL and its secret language version, BLAZIR. In such cases, the phoneme that was closest to the original but which was still permissible in the new phonetic context was taken as the "correct" form.

Error analysis. There were three major types of errors: 1) failure to transform a target phoneme (e.g., LIVER⟶RIVER rather than RIVEL); 2) phonemic changes in nontarget phonemes (e.g., OFFER⟶OHFUL rather than AWFUL); 3) intrusions (e.g., BRAMBLE⟶BLAMBLER). A composite error score for each subject was obtained by summing the number of all types of errors. The average number of errors was 21 per subject.

Output tempo. Another interesting aspect of the secret language responses was output tempo, which is illustrated in Figure 1. At the top of the display is a diagrammatic representation of a stimulus item. The time line moves from left to right while the hatch marks indicate that audible sound is being produced. While this is a very crude representation of speech, it does serve the present analysis. There were two general kinds of output. 1) Global response: there was only a brief pause following stimulus presentation; then the entire item was uttered in transformed version. 2) Sequential response: there was a fairly long pause after the stimulus item; then the subject gave part of the response, paused, gave some more, sometimes paused again, and finally finished his response. The following taped passages illustrate these two forms of response. First, a subject who characteristically gave global responses:

STIMULUS	SUBJECT T.C.
LEVER	REVEL
BRAMBLE	BLAMBER
LAVISH	RAVISH
MIDDLE FINGER	MIDDER FINGLE

This subject barely paused after the stimulus item, then produced the whole item in secret language form. Next is a subject who characteristically gave sequential responses:

STIMULUS	SUBJECT R.C.
PICKLE	...PICK...KER
LIKELY	...RIKE...RY
BRAMBLE	...BLAM...BER
MIDDLE FINGER	...MIT...TER...FING...GLE

This subject paused for a fairly long time before beginning his response and gave small response units interspersed with additional pauses.

Orthographic influences. Some subjects seemed to stay within the auditory mode: given the sound units of the stimulus, they changed the appropriate units and gave the resulting phoneme string as their response. Others seemed to convert the input phonemes into an orthographic representation before any transformations were made. Figure 2 illustrates the two approaches for the stimulus word LITTLE. The phonemic representation of the stimulus is shown on the left side of the display; the "T" sound is actually a flapped /t/ which sounds like

Figure 1

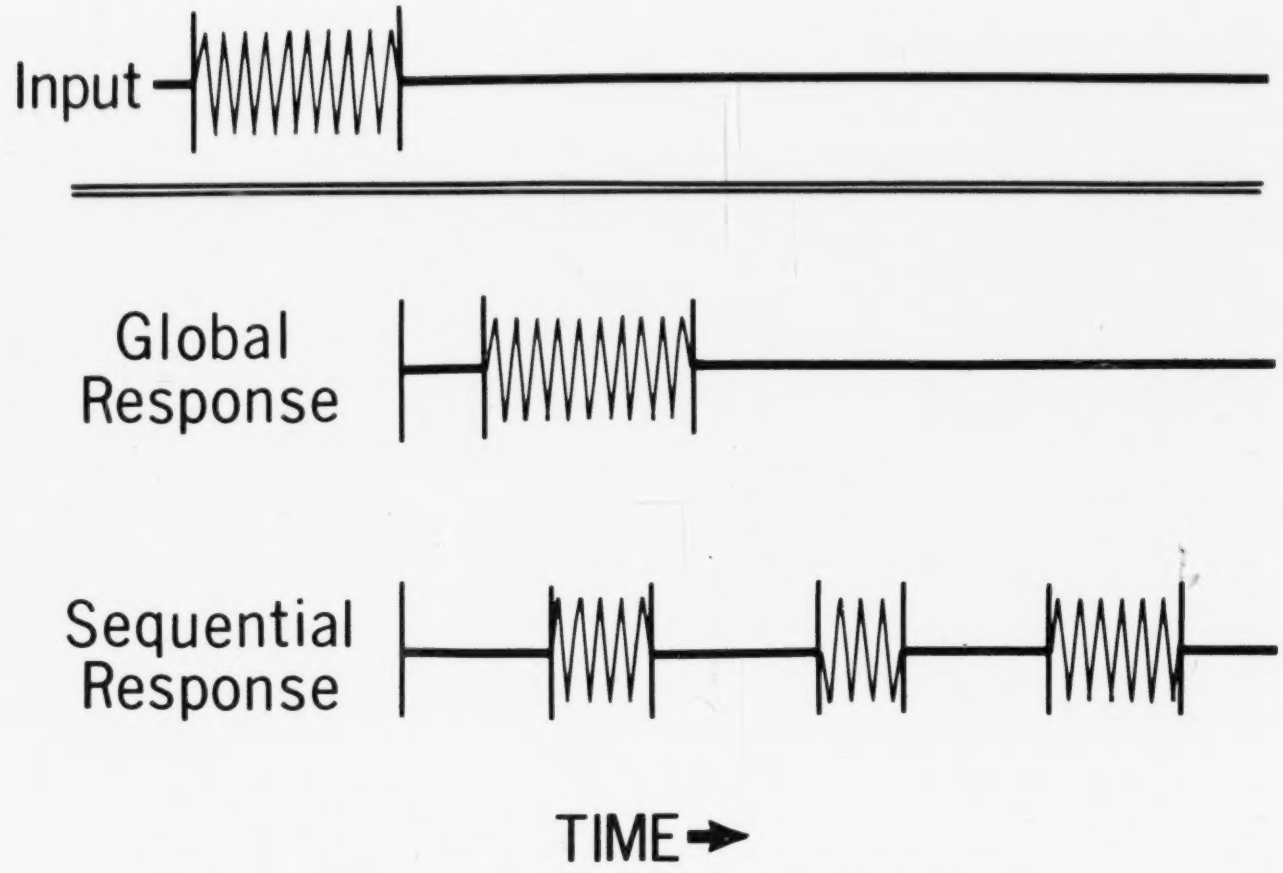


Figure 1: Schematic representation of global vs. sequential output tempos.

Figure 2

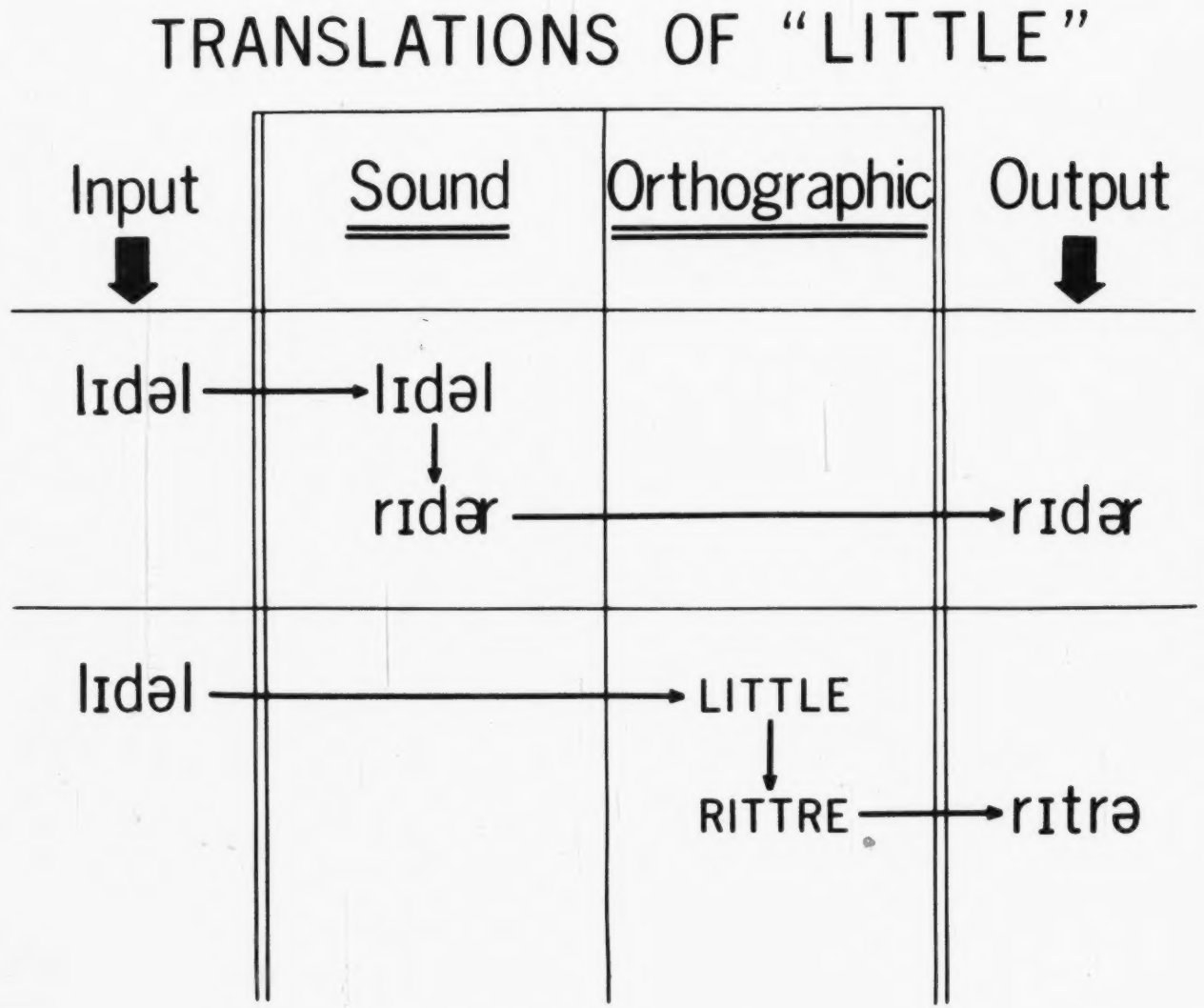


Figure 2: Two translations of a stimulus word: sound only vs. the addition of orthographic influences.

/d/. One way to transform LITTLE into R-L language is RIDDER. This transformation stays entirely within the sound mode: the initial /l/ is transformed to /r/, the final /l/⟶/r/, such that the output is /rIdər/. Other subjects give /rItrə/ as their response. Given the input /lIdəl/, they appear to transform it into an orthographic mode of representation: L-I-T-T-L-E. Transformations are then made on letters; the letter "L" is replaced by "R" in two locations, yielding R-I-T-T-R-E. This new orthographic representation is then turned back into a sound representation, such that the subject says /rItrə/ rather than /rIdər/.

Individual Differences

Earlier we pointed out the wide range of individual differences in success of making /r/⟷/l/ transformations. Were there systematic individual differences in the present experiment? In order to answer this question, a brief discussion of systematic individual differences in another paradigm is necessary.

Studies of dichotic fusion. Large and systematic individual differences have been obtained in the dichotic fusion situation (Day, 1969). Briefly, a different message is presented to each ear at the same time over earphones and subjects are asked to report 'what they heard.' The dichotic items are of the general form BANKET/LANKET. On a substantial proportion of trials, fusions occur: subjects report hearing BLANKET. The extent to which a population of subjects fuses is of primary interest for the present discussion. Each subject is given a fusion score, which is simply the percent of trials on which he gave a fused response such as BLANKET. We then look at the number of subjects who achieved the full range of scores. Ordinarily, in most psychological tests, a normal distribution is obtained, whether the scores represent percent correct, number of trials to criterion, or a variety of other measures. However, dichotic fusion scores are not normally distributed, but are instead bimodal in nature: some subjects fuse most of the time, while others rarely fuse. Since the first bimodal distribution of fusion scores was obtained (Day, 1969), the effect has been replicated many times over several hundred subjects.

Another task, which uses the same dichotic tapes, asks the same subjects to determine which phoneme begins first on every trial. For an item like BANKET/LANKET, the /b/ begins first by a short interval (e.g., 25-150 msec) on half the trials, while the /l/ begins first by the same intervals on the remaining trials. The high fusers report hearing the /b/ first on most of the trials. Note that in English /bl-/ can occur in initial position but /lb-/ cannot. Therefore the high fusers are reflecting the phonological rules of English; they are not reflecting the true stimulus conditions. These subjects have been termed "language-bound" since they are bound by the facts of their language so that they cannot judge the target stimulus events accurately. When the low fusers are asked to judge temporal order they are highly accurate, whether the /b/ or the /l/ began first. These subjects have been termed "stimulus bound" since they are accurate judges of the target stimulus events.

Current secret language experiment. Some of the subjects in the present experiment also took the dichotic fusion tests. Eleven were classified as language-bound and eleven as stimulus-bound. There were clear differences in secret language facility between the two groups. The data given below are for the word translation task; a similar pattern of results was obtained on the translations of "Mary had a little lamb."

The language-bound subjects made almost twice as many errors: their composite error score was 23 errors per person as compared with 13 for the stimulus-bound subjects. Most of the differences between the two groups occurred on W⟶NW items where the composite error scores were 20 and 10, respectively.

Over all items, language-bound subjects gave more global than sequential responses, while stimulus-bound subjects gave an equal number of both types of output tempo. Again, the type of item made a difference. For language-bound subjects, the percent of responses that were global vs. sequential was 80% vs. 20% for W⟶W items and 48% vs. 52% for W⟶NW items. For stimulus-bound subjects, these figures were 60% vs. 40% for W⟶W items and 37% vs. 63% for W⟶NW items. Thus, while both groups of subjects gave more global responses on W⟶W items, it was only the stimulus-bound subjects who gave more sequential responses on W⟶NW items. It might be of interest to teach language-bound subjects to use a sequential strategy on W⟶NW items. While their performance may improve somewhat, they still may do poorly, as illustrated by the attempts described above to get Subject D.Q. to transform BRAMBLE syllable-by-syllable.

Language-bound subjects gave four times as many responses that reflected orthographic influences. They were thus less able to make transformations solely at the sound level, independent of written structure. These are the same subjects who had difficulty in making temporal order judgments in the dichotic fusion situation independent of phonological rules. The stimulus-bound subjects were not misled by orthographic conventions in the secret language experiment, nor were they misled by phonological constraints in dichotic fusion tests.

Discussion. Perhaps Saussure's (1915) notion of la langue (language) and la parole (speech) is helpful in understanding the two groups of subjects. Stimulus-bound subjects are able to track the "speech" end, that is, the actual performance aspects of an utterance. Language-bound subjects, on the other hand, perceive an utterance through "language," that is, through the abstract structure of their language.

The secret language experiment can be used for at least two types of research. 1) "Psychological reality." Secret language rules can be added at various levels of linguistic analysis, for example, the phonetic, phonological, syllabic, syntactic, or semantic levels. The relative ease with which subjects can make these transformations may reflect the extent to which each level or type of rule is psychologically "real." 2) Individual differences. The secret language experiment is well adapted to studying individual differences in language ability. The ease with which individuals can operate on linguistic structure may well have predictive value for foreign language learning.

In conclusion, I introduce you to the secret language experiment as a new tool for studying psycholinguistic phenomena. It also happens to be an enjoyable experience, both for the subject and the experimenter.

REFERENCES

Alexis, Gerson. (1970) Le parler bolite. In Lecture en Anthropologie Haïtienne. (Port-au-Prince: Presses Nationales d'Haiti) 203-207.

Conklin, Harold C. (1956) Tagalog speech disguise. Language 32, 136-139.

Day, Ruth S. (1969) Temporal order judgments in speech: Are individuals language-bound or stimulus-bound? Paper presented at the 9th meeting of the Psychonomic Society, St. Louis, Mo., November. [Also in Haskins Laboratories Status Report on Speech Research (1970) SR-21/22, 71-87.]

Price, Richard and Sally Price. (in press) Secret play languages in Saramaka: Linguistic disguise in a Caribbean creole. In Speech Play on Display, ed. by Barbara Kirshenblatt-Gimblett. (The Hague: Mouton).

Saussure, Ferdinand de. (1915) Course in General Linguistics. (Reprinted by McGraw-Hill, 1966.)

Scherzer, Joel. (1970) Talking backwards in Cuna: The sociological reality of phonological descriptions. Southwest. J. Anthropol. 26, 343-353.

Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes

M. F. Dorman[+] and Donna S. Geffner[++]

ABSTRACT

Six-year-old black and white Ss from low and middle socioeconomic classes (SEC) were presented a dichotic listening task composed of syllable pairs. All groups evidenced a significant right-ear advantage (REA) at recall. The magnitude of the REA did not differ as a function of race or SEC. The magnitude of the REA averaged over all Ss was similar to that of adults.

On dichotic listening tests with adults, when verbal stimuli are presented simultaneously to the left and right ears, the stimuli presented to the right ear are recalled better than those presented to the left ear (Kimura, 1961a; Bryden, 1963; Broadbent and Gregory, 1964; Curry and Rutherford, 1967; Shankweiler and Studdert-Kennedy, 1967). This right-ear advantage (REA) presumably reflects the functional prepotency of the contralateral auditory pathways and the left hemisphere's specialization for the perception of speech (Kimura, 1961b; Milner, Taylor, and Sperry, 1968; Studdert-Kennedy and Shankweiler, 1970).

In children the REA appears to vary as a function of age, sex, and socioeconomic class (SEC) background. Kimura (1967) found that low SEC females and high SEC males and females evidence a REA at age five, whereas low SEC males do not evidence a REA until age six. Recently Geffner and Hochberg (1971) have reported a large (age) X (SEC) interaction in the development of the REA. Four- to seven-year-old Ss from both low and middle SEC backgrounds were presented a dichotic digits task (cf. Kimura, 1963). The middle SEC Ss evidenced a REA at all ages. The low SEC Ss did not evidence a REA until age seven. These data led Geffner and Hochberg to speculate that children from low SEC backgrounds may not develop left-hemisphere specialization for speech at the same rate as children from more privileged SEC backgrounds.

The Geffner and Hochberg data are very striking for they suggest that cortical lateralization of function, which has been thought to be maturationally

---

[+]Haskins Laboratories, New Haven, Conn., and Herbert H. Lehman College of the City University of New York.

[++]Herbert H. Lehman College of the City University of New York.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

determined (Lenneberg, 1967), may be radically slowed down by environmental deficiencies during development. The nature of the environmental conditions which determined the performance of the low SEC Ss is, however, not at all clear. One nonenvironmental variable which may have affected the outcome of the Geffner and Hochberg study was the large proportion of black children in the low SEC group. Conceivably the delayed lateralization of speech found for the low SEC population may have been a racial effect interacting with socioenvironmental variables.

To clarify the effects of race and SEC as determinates of the REA in children, in the present study, six-year-old black and white children from both low and middle SEC backgrounds were presented a dichotic syllable test (cf. Studdert-Kennedy and Shankweiler, 1970). To minimize a possible source of experimental bias, the groups of children were tested by an examiner of their own race.

Another purpose of the present study was to compare the REA of six-year-old children with previously collected data on the REA in adults (Studdert-Kennedy and Shankweiler, 1972). If, as Lenneberg (1967) has suggested, cortical lateralization of function is not complete until approximately puberty, then we may expect that six-year-old Ss would evidence a smaller REA than adults. If, however, lateralization is complete by age five (cf. Krashen and Harshman, 1972) then we may expect that the magnitude of the REA in six-year-old children and adults would be similar.

METHOD

Subjects. The Ss were 52 six-year-old children (C.A. 6.0-6.8): 26 white Ss, 13 each from low and middle SEC; 26 black Ss, 13 each from low and middle SEC. SEC was determined by Hollingshead's Two Factor Index of Social Position (Hollingshead, 1965) which takes into account the parents' educational level and occupational status. All Ss were right-handed (handedness tasks are detailed in the Procedure) and had normal hearing with no known perceptual, neurological, speech, or language deficit. Children with a bilingual background were not selected. The Ss were matched as well as possible by class placement and performance. Because intelligence quotients are not available in New York City public schools, the authors obtained all information pertaining to the parents' occupation and educational level, the home evnironment of the Ss, and classroom performance from the classroom teacher, principal, guidance counselor, or parent.

Apparatus. The stimuli were reproduced on a Roberts 1920 stereo tape recorder via matched TDH-39 headphones. The output of each tape channel was calibrated and monitored by a Hewlett-Packard voltmeter. A 1000 Hz tone on both channels of the test tape was used as a calibration signal. Audiometric threshold tests were administered on a Maico MA-10 portable audiometer calibrated to ISO standards.

Preparation of stimuli. With the aid of the Haskins Laboratories' computer-controlled parallel resonance speech synthesizer the stop consonant-vowel syllables /ba, da, ga, pa, ta, ka/ were generated. Each stimulus was composed of three formants, and was 300 msec in duration. Under computer control these six stimuli were then combined into the 15 possible pairs (no stimulus was paired with itself) and were recorded dichotically in a fully balanced order onto

magnetic tape. The resulting tape contained 60 stimulus pairs with each member of a pair occurring twice on each channel. The interstimulus interval was 4 sec.

Procedure. Each S was tested in a quiet room, most often the school nurse's. All Ss were first given an audiometric threshold test. Hearing level at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz was assessed. If the hearing level between the two ears differed by 10 db or more for two of the test frequencies, the S was excused from further testing. Handedness was determined by asking the Ss to perform three manual motor tasks: throwing a ball, cutting with scissors, and drawing a circle. Any S who did not perform all three tasks with his right hand was not tested further.

After the preliminary examination, the Ss were presented binaurally three repetitions of the syllables /ba, da, ga, pa, ta, ka/. The Ss were instructed to listen with both ears and report the syllable heard. Any S unable to repeat the six syllables after the third repetition of the list was excused from further testing. Next, the Ss were presented three dichotic practice trials. Again the Ss were instructed to listen with both ears and report the syllable heard. (Since the Ss were not told there were two different stimuli on these and the following dichotic trials, only one response was elicited.) The Ss were told that these sounds would sound "funny" but to continue reporting them as before. The Ss were then presented the 60-item test sequence, followed by a brief rest, then the 60-item test again. To control for possible channel effects, the headphones were reversed after each 60-item test. The black Ss were tested by a black student assistant, while the white Ss were tested by a white examiner.

Results

Each S's performance was scored in terms of the metric $\frac{R-L}{R+L} \times 100$ where R is the total number of items correctly reported from the right ear and L is the total number of items correctly reported from the left ear (for a discussion of this scoring technique see Studdert-Kennedy and Shankweiler, 1970). The mean score for each of the groups subcategorized by sex is shown in Table 1.

TABLE 1: Magnitude of the REA for all groups in terms of $\frac{R-L}{R+L} \times 100$.

		RACE: White	RACE: Black
SEC: Middle		Males (n = 5) = 15.19	Males (n = 5) = 9.69
		Females (n = 8) = 14.53	Females (n = 8) = 16.72
		Average (n = 13) = 14.86	Average (n = 13) = 12.17
SEC: Low		Males (n = 4) = 0.02	Males (n = 11) = 10.15
		Females (n = 9) = 9.54	Females (n = 2) = 10.22
		Average (n = 13) = 7.14	Average (n = 13) = 10.16

Each (race) X (SEC) group's mean score was evaluated by *t* tests for correlated samples against the hypothesis that there was no difference in accuracy of report between the left and right ears. As shown in Table 2, all groups evidenced a significant REA.

TABLE 2: Mean number of syllables correctly reported for each ear.

Group	N	Left	Right	*t*
black low SEC	13	35.00	42.92	3.50**
black middle SEC	13	34.76	44.86	2.26*
white low SEC	13	41.46	47.87	2.86*
white middle SEC	13	33.61	48.38	2.19*

*p < .05
**p < .01

To determine whether the magnitude of the REA differed as a function of race or SEC, the individual scores were collapsed over sex into an analysis of variance with race and SEC as treatment variables. Neither the race ($F\ 1|48 = .021$, $p > .05$) nor the SEC ($F\ 1|48 = 1.009 > .05$) main effect was significant. The (race) X (SEC) interaction, suggested by the reduced REA for the low SEC white Ss, was not significant ($F\ 1|48 = .405$, $p > .05$).

Statistical analysis of male-female differences was not attempted because of the generally small sample size, especially for males (n = 2) in the black low SEC condition. Because two of the four male white low SEC Ss evidenced rather large left-ear advantages, no overall ear advantage occurred for this group. There is no reason to suspect that the obtained REA is representative of the entire white male low SEC population.

DISCUSSION

Racial Factors in REA

No difference in magnitude of the REA was found between black and white Ss. A similar outcome is reported by Sadick (in preparation). Black and white five- and seven-year-old Ss were presented a dichotic syllable test similar to that used in the present study. At both the five- and seven-year levels, both black and white Ss evidenced a REA. The magnitude of the REA did not differ between the groups. We may then tentatively conclude that the rate of cerebral lateralization of function does not vary as a function of racial origin. This conclusion is, of course, limited to the racial groups and SEC environments studied.

SEC Factors in REA

The presence of a significant REA in the low SEC groups, although conflicting with the outcome of Geffner and Hochberg (1971), is consistent with the

outcome of a recent study by Knox and Kimura (1970). These investigators assessed cerebral lateralization for speech and nonspeech sounds in five- to eight-year-old low SEC Ss. The Ss were presented both dichotic digits and dichotic environmental sounds. In the digit condition, all age-sex groups evidenced a REA. Thus, in Kimura's several studies of the REA in five-year-old low SEC Ss (Kimura, 1967; Knox and Kimura, 1969) both males and females have evidenced REAs, although males less consistently than females.

These data, paired with the significant REA in the six-year-old black and white low SEC Ss found in the present study, suggest that at least some, perhaps the majority, of low SEC Ss achieve left-hemisphere specialization for speech at the same rate as higher SEC Ss.

One possible explanation for the difference in outcome between the present study and that of Geffner and Hochberg (1971) is that the low SEC Ss examined by the latter investigators may have been raised in more deprived environments than those of the present study. Geffner and Hochberg argued that abnormal rearing conditions may have resulted in a retarded rate of cerebral lateralization of function. However, an alternative and somewhat less radical explanation is that abnormal rearing conditions engender Ss who function at very low cognitive and motivational levels. Such Ss might perform "indifferently" on a relatively complex task like dichotic digits, especially when tested by someone not of their own race. On this view it would be expected that Geffner and Hochberg's four-, five-, and six-year-old low SEC Ss would evidence very low overall performance levels on the digits task. An analysis of the Geffner and Hochberg data has indicated that, indeed, the low SEC Ss reported only 53% of the total possible digits, while the middle SEC Ss reported 62%. Thus the low SEC Ss evidenced a significantly lower performance level ($t_{154} = 3.98, p < .001$). This outcome suggests that the absence of a REA in the low SEC Ss may have been a "floor effect" (cf. Halwes, 1969) resulting from task difficulty and motivational variables. To choose between the alternative explanations for the absence of a REA, it would appear necessary to present four- and five-year-old very low SEC Ss with a relatively simple dichotic test (e.g., dichotic syllables) in a situation which would maximize the Ss' motivational level. Until such data have been collected, the effect of rearing conditions on the rate of cerebral lateralization of function remains unclear.

Development of the REA

Lenneberg (1967) has suggested that the end of the critical period for language acquisition and the terminus of cerebral lateralization of function occurs at approximately puberty. On this view, we may suspect that the magnitude of the REA would increase until puberty.

The data from several experiments indicate, however, that the REA may not systematically increase in magnitude between age five and adulthood. In the present study of six-year-olds the magnitude of the REA averaged over groups was 11.08. This magnitude REA is well within the range of REAs found for samples of the adult population tested with a dichotic syllable procedure. Dorman and Porter (1971) found a REA (n = 10) of 12.0 while Studdert-Kennedy and Shankweiler (1972) report a REA (n = 30) of 10.0. The proportion of Ss (.21) with a left-ear advantage is also similar to that found for adult populations (Studdert-Kennedy, personal communication).

In a developmental study of the REA, Berlin, Lowe-Bell, Hughes, and Berlin (1972) administered a dichotic syllable test to male and female Ss 5 to 13 years old. All age-sex groups evidenced a REA. Although the total number of correct responses increased with age, the magnitude of the REA did not systematically increase with age. From these data and those collected from adults using the same test, Berlin et al. (1972) concluded that the REA may develop fully by age five. Krashen and Harshman (1972) have also reached this conclusion after a re-analysis of earlier dichotic listening data (Geffner and Hochberg, 1971; Kimura, 1963; Knox and Kimura, 1970), taking into account changes in guessing strategy with increased age. The "lateralization by age five" hypothesis also appears consistent with clinical data on language disturbance following brain injury (Krashen and Harshman, 1972).

Finally, a number of studies indicate that the cortical mechanisms underlying the perception of speech may be lateralized in very young children, perhaps even infants. Yeni-Komshian (personal communication) has found a REA in some three-, four-, and five-year-old children (also see Nagafuchi, 1970), while Molfese (1972), using an auditory evoked response technique, has reported larger left-hemisphere than right-hemisphere responses to speech signals in infants. Taken together, the studies cited above suggest that cortical specialization for speech perception may be present in very young children and may be complete by ages five to six.

REFERENCES

Berlin, C., S. Lowe-Bell, L. Hughes, and H. Berlin. (1972) Dichotic right-ear advantage in males and females--ages 5-13. Paper presented at the 84th meeting of the Acoustical Society of America, Miami Beach, Fla., November.

Broadbent, D. and M. Gregory. (1964) Accuracy of recognition for speech presented to the right and left ears. Quart. J. Exp. Psychol. 16, 359-360.

Bryden, M. (1963) Ear preference in auditory perception. J. Exp. Psychol. 65, 103-105.

Curry, F. and D. R. Rutherford. (1967) Recognition and recall of dichotically presented verbal stimuli by right- and left-handed persons. Neuropsychologia 5, 119-126.

Dorman, M. and R. Porter. (1971) Hemispheric specialization for speech in stutterers and nonstutterers. Unpublished manuscript.

Geffner, D. and I. Hochberg. (1971) Ear laterality performance of children from low and middle socioeconomic levels on a verbal dichotic listening task. Cortex 2, 193-203.

Halwes, T. (1969) Effects of dichotic fusion on the perception of speech. Unpublished Ph.D. thesis, University of Minnesota (Psychology). (Issued as Supplement to Haskins Laboratories Status Report on Speech Research, September.)

Hollingshead, A. (1965) Two Factor Index of Social Position. (New Haven, Conn.: Yale Station).

Kimura, D. (1961a) Cerebral dominance and the perception of verbal stimuli. Canad. J. Psychol. 15, 166-171.

Kimura, D. (1961b) Some effects of temporal-lobe damage on auditory perception. Canad. J. Psychol. 15, 156-165.

Kimura, D. (1963) Speech lateralization in young children as determined by an auditory test. J. Comp. Physiol. Psychol. 56, 899-902.

Kimura, D. (1967) Functional asymmetry of the brain in dichotic listening. Cortex 3, 163-178.

Knox, C. and D. Kimura. (1970) Cerebral processing of nonverbal sounds in boys and girls. Neuropsychologia 8, 227-237.

Krashen, S. and R. Harshman. (1972) Lateralization and the critical period. J. Acoust. Soc. Amer. 52, 174(A).

Lenneberg, E. (1967) Biological Foundations of Language. (New York: Wiley).

Milner, B., L. Taylor, and R. Sperry. (1968) Lateralized suppression of dichotically presented digits after commissural section in man. Science 161, 184-184.

Molfese, D. (1972) Cerebral asymmetry in infants, children, and adults: Auditory evoked responses to speech and noise stimuli. Unpublished Ph.D. thesis, Pennsylvania State University (Psychology).

Nagafuchi, M. (1970) Development of dichotic and monaural hearing abilities in young children. Acta Otolaryng. 6, 409-414.

Sadick, T. (in preparation) Doctoral dissertation, Department of Biobehavioral Sciences, University of Connecticut, Storrs.

Shankweiler, D. and M. Studdert-Kennedy. (1967) Identification of consonants and vowels presented to left and right ears. Quart. J. Exp. Psychol. 19, 59-63.

Studdert-Kennedy, M. and D. Shankweiler. (1970) Hemispheric specialization for speech perception. J. Acoust. Soc. Amer. 48, 579-594.

Studdert-Kennedy, M. and D. Shankweiler. (1972) A continuum of cerebral dominance for speech perception? Haskins Laboratories Status Report on Speech Research SR-31/32.

Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech

Gloria Jones Borden,[+] Katherine S. Harris,[++] and William Oliver[+++]

The effects of bilateral mandibular nerve blocks on speech were judged by listeners and by transcribers. Seven adult male speakers repeated under normal and nerve-block conditions 66 sentences heavily weighted with consonant clusters known from pilot studies to be vulnerable to nerve-block distortion. Listener judgments of the speech revealed large magnitudes of subject variance. Although all subjects reported loss of sensation, the effects on speech ranged from completely unaffected to markedly affected. Distortions were noted by narrow phonetic transcription in 23% of the data, most prominently in /s/ clusters.

The question of whether skilled speech is an open loop system requiring little or no feedback from the periphery, or a closed loop system requiring sensory information to control the production is a provocative topic and basic to our understanding of speech patterning. One feedback channel, that of sensation from the oral cavity, can be studied by examination of the effects of sensory deprivation. One approach is to examine the speech of subjects in which an oral sensory deficit is pathological, but this method yields contradictory conclusions (Chase, 1967; McDonald and Aungst, 1970). It is difficult to obtain specific information on the relationship between oral sensation and speech from clinical cases, due to the multiplicity of handicaps.

[+]Haskins Laboratories, New Haven, Conn., and City College of the City University of New York.

[++]Haskins Laboratories, New Haven, Conn., and the Graduate Division of the City University of New York.

[+++]University of Pennsylvania School of Dentistry, Philadelphia, Pa.

Acknowledgment: This article summarizes a portion of a doctoral dissertation by the first author completed at the Graduate Division of the City University of New York under the direction of Katherine S. Harris (1971). The authors gratefully acknowledge the assistance of Harry Levitt and Stewart Lawrence at City University of New York, Lorraine H. Russell at Temple University, Philadelphia, Pa., and Tom Gay at the University of Connecticut Health Center, Farmington. This research was supported in part by a grant from the National Institute of Dental Research to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

A potentially productive way of studying the relationship between sensory feedback from the oral area and articulation is to interrupt feedback by blocking the trigeminal nerve in normal speakers.

It is frequently observed that after dental procedures involving nerve blocks there is often a disturbance of clearly articulated speech until the effect of the anesthesia has disappeared. It is understandable, therefore, that investigators interested in afferent control of speech should block the sensory nerves of normal speakers with anesthesia in order to study the relationship between feedback from the oral area and articulation of speech. Presumably all feedback channels are used to acquire language: audition, taction, and proprioception. Do normal adult speakers need to depend upon these feedback possibilities during ongoing speech, and to what degree or under what circumstances does each channel play a role? McCroskey (1958) was the first to report that blocking oral sensation with mandibular and intraorbital injections of anesthesia had an adverse effect on articulation. Substitution and distortion errors were reported (McCroskey, Corley, and Jackson, 1959). Ringel and Steer (1963) confirmed the findings of McCroskey. It was assumed that the reason for the articulatory deterioration was the interruption of a closed loop control system. Locke (1968) questioned the technique, as it might have both motor and sensory effects, but Schliesser and Coleman (1968) reported complete elimination of tongue sensation as tested by oral stereognosis measures after mandibular blocks and the application of a topical anesthetic to the anterior palate. They also reported very little, if any, interference with the motor control needed to lateralize the tongue or to perform diadochokinetic tasks. Several investigators interested by the McCroskey study and the Ringel and Steer study attempted to specify further the effects of the nerve block. Work was done on this subject somewhat concurrently by Gammon, Smith, Daniloff, and Kim (1971), by Scott (1970), and by the authors. Gammon and colleagues found a 20% rate of misarticulation with anesthesia. Errors were more prominent in the labial and alveolar regions, with fricatives and affricates especially distorted. Scott noted that sibilants were less closely produced and other phonemes were retracted but maintained the intended manner of articulation. The stimuli used were 24 spondee words.

The purpose of this study was to investigate further the distortion of phonemes vulnerable to nerve block. Is the effect upon speech slight or severe? Does it affect subjects similarly? What phonemes are distorted?

METHOD

Two pilot studies, the first using a consonant-vowel-consonant (CVC) balanced list and the second using words containing fricatives, revealed that speech deterioration under nerve block was in many cases evident only in rapid, connected speech. Sixty-five sentences were, therefore, constructed for the final study (Borden, 1971). The subjects were seven university students, all normal speakers of standard English. The recording was done in a quiet interior room at the University of Pennsylvania School of Dentistry. Each subject had two sessions. The conditions of normal and nerve block were rotated as far as possible. In each session the subject repeated the sentences after listening to a recorded speaker heard through earphones from a second tape recorder. The anesthesia was administered by a dentist using the standard dental technique for producing a mandibular block (Cook-Waite Labs, Inc., 1971). The puncture was made at the apex of the pterygomandibular triangle which is about 7 mm above the occlusal

surface of the teeth. Half of the solution of 2% lidocaine was deposited halfway back toward the wall of the mandibular sulcus. This usually anesthetizes the lingual nerve. When the needle reached the ramus, the rest of the solution was deposited around the inferior alveolar nerve. The method of injection is schematized in Figure 1. The amount of anesthesia was 1.5 cc of solution on each side. When subjects reported loss of sensation to the dentist's probes of the tongue, taping began. In some instances, an additional 1.5 cc was injected if needed. The anesthesia lasted for the session which took little more than one-half hour.

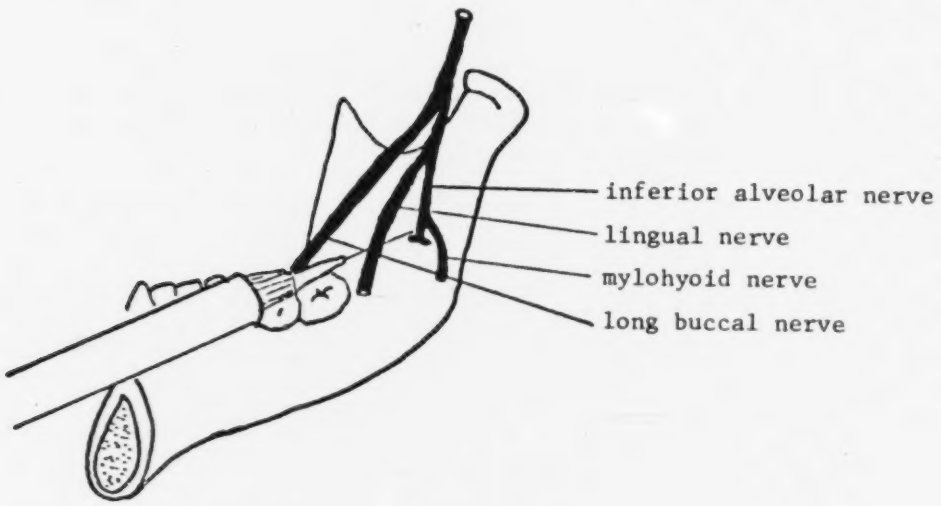


Figure 1: Inner surface of ramus with needle in the right mandibular sulcus.

Listening Test

Thirty-eight utterance samples in the form of phrases were extracted from the recorded material to construct a listening test. This test was used for listener judgments of speech deterioration and for narrow phonetic transcriptions by transcribers. The test was heavily weighted with utterances that from pilot studies were found to be most vulnerable to the nerve block.

The tapes of each subject were presented to a group of listeners. Utterance samples of both conditions had been spliced into matched pairs, randomized, and separated by one second of silence between each one of a pair and four seconds of silence between pairs of utterances. The listeners were 16 university students instructed to check a if the first example of each pair seemed more deteriorated, or check b if the second example seemed more deteriorated. The incorrect responses (checking normal condition as deteriorated) were counted and tabulated according to speaker and according to utterance. Correspondence between those utterances sampled during nerve-block conditions and the listener response "more deteriorated" served as an index of nerve-block influence.

Phonetic Transcriptions

Two experienced transcribers made narrow phonetic transcriptions of the listening test tapes. The transcribers worked on material and speakers not used in the study to standardize their phonetic system. It was decided that the direction of the distortions should be indicated whenever possible. For example, if the /s/ sounded somewhat like /θ/, the transcription would be /s$^{\theta}$/, whereas if it was more toward /ʃ/, it would be transcribed /s$^{\int}$/. If the /s/ was slurred but in an undetermined direction, the indication was /s̱/.

RESULTS

Listener Judgments

Incorrect listener responses were tabulated according to speakers and according to utterances. Two analyses of variance were conducted to investigate variation among listeners and among speakers and further, the variation among utterances according to listeners (Borden, 1971).

It was found that there was no significant difference among listeners. Thus, the listeners were apparently using the same criteria in their judgments.

The variation among utterances according to speaker was significant at the .05 level, indicating marginal significance. One speaker who evidenced no speech distortions under nerve block was removed for this analysis. The total possible "incorrect" listener responses for each utterance was 96 (6 speakers as heard by 16 listeners) of which 48 would be expected by chance, even without a nerve block. In general the single consonants deteriorate less than the clusters since listeners have more trouble identifying the block condition.

The variations among speakers was found to be highly significant as judged by listeners. Since there were 38 utterances and 16 listeners, there were 608 possible incorrect listener responses for each speaker, 304 expected by chance. It can be seen in Table 1 that the nerve block had no effect on speaker B (315 incorrect responses) as determined by listener judgment. Speaker C, in contrast,

TABLE 1: Incorrect listener responses according to speaker.

Speakers	B	A	E	F	G	D	C
Total	315	246	228	222	218	193	96
% Utterances	50	40	38	37	36	32	15

was most affected, as the listeners made relatively few errors of judgment (96) between the normal and the nerve-block utterances. Speakers, then, varied considerably in their performance under nerve block as judged by listeners. Even when the speaker with no perceptible effect on his speech was removed for the second analysis, a significant variation among the remaining 6 speakers was found

at the .01 level of confidence. The extent of this variation was surprising to the experimenters, as there had been no previous mention of interspeaker variance in levels of deterioration due to oral anesthesia.

Transcriber Judgments

The transcribers made the transcriptions independently. Transcriber agreement was quite high. For the 228 utterances transcribed, transcribers agreed that there was no effect in 67% of the data. There was agreement both that there was a distortion effect and on the nature of that distortion in another 20% of the data, bringing the transcriber agreement up to 87%. Of the total number of utterances 3% had transcriber agreement that there was a deviation, but the direction or place of the distortion in the utterance was judged differently. For the final 10% of the data, one transcriber heard differences which the other did not consider to be distortions.

To determine if the transcribers were making judgments on utterances similar to the judgments made by the 16 listeners, the utterances were ranked according to transcriber judgments of deterioration to test the correlation of that ranking with the utterances ranked according to listener errors. The utterances were given scores to indicate their relative degree of distortion as interpreted by the two transcribers. An utterance received a score of zero if there was no difference noted by either transcriber between the normal and nerve-block condition in any speaker. A score of 1/2 indicated that a difference was noted by one transcriber in one speaker, and 3/4 indicated that a difference was noted by one transcriber in 2 speakers. Scores of 1, 2, 3, or 4 were assigned if there was transcriber agreement that there was a distortion in 1, 2, 3, or 4 speakers respectively. After each utterance was assigned a score, the utterances were ranked from the most affected by the block to the least affected. Table 2 shows the key words removed from their embedding phrases as ranked by transcribers and by listeners. Using Spearman's Rank Correlation, the ranking of utterances given by the transcribers correlated significantly at the .01 level of significance with the ranking of utterances given by the 16 listeners.

The phonemes transcribed as distorted under nerve block were /tʃ/, /dʒ/, /s/, /z/, /ʃ/, /t/, and /l/. All of the /s/ two-consonant clusters were distorted, especially /st/. Among the /s/ three-consonant clusters only the final /kst/ remained undistorted by the block. The /s/ was the distorted portion of the cluster in all cases, with additional distortion on /r/ in two utterances with /spr/ and /skr/ clusters. There were no errors transcribed for the labials, the velars, the labiodentals, or for /d/ or /n/.

All of the errors noted by the transcribers were errors of place. The errors were never sufficiently deviant to cross phoneme boundaries. The most prominent distortion was for the /s/ to deviate toward /ʃ/. In all cases the distortion seems to be the result of the tongue failing to reach target position or of target precision.

Speaker Variation

Transcriber judgments according to speaker indicate, as did the listeners, that the speakers varied widely in the degree of speech deterioration evidenced in the sample utterances. Listeners and transcribers agreed that speaker C was

TABLE 2: Rank correlation between transcribers and listeners.

Utterance	Transcription Score	Transcriber Rank	Listener Rank
spring	4.5	1.5	2
stars	4.5	1.5	1
scissors	4	3.5	12.5
school	4	3.5	15
squirrel	3.5	5.5	10
watching	3.5	5.5	23
spider	3	8	3
whiskers	3	8	18
scratch	3	8	6
letters	2.5	10	20
mouse	2	11.5	28.5
string	2	11.5	7.5
dishers	1.75	14	23
snowballs	1.75	14	7.5
giraffe	1.75	14	10
blocks	1.5	18	26
brushing	1.5	18	27
bicycles	1.5	18	28.5
grapes	1.5	18	16
smoke	1.5	18	4.5
sweeping	1	23	4.5
sleeping	1	23	18
it's	1	23	10
kids	1	23	25
splashing	1	23	21
telephone	.75	26	12.5
knife	.5	28	35
swinging	.5	28	23
shaving	.5	28	31.5
table	0	34	14
pajamas	0	34	18
girl	0	34	31.5
bird	0	34	34
fixed	0	34	31.5
birthday	0	34	31.5
mother	0	34	37.5
cans	0	34	36
peanut	0	34	37.5

the most affected, and that speaker B was not affected. As demonstrated in Table 3, speakers C and D were both affected, speakers F and G somewhat less affected, and speakers E and A were judged to have very little deterioration of speech.

TABLE 3: Speaker variation as judged by transcriber agreement of no distortion.

Speaker	B	E	A	G	F	D	C
% Utterances Judged Not Distorted	100	89	82	74	58	55	45

DISCUSSION

Several results emerged from this perceptual study of the effects of bilateral mandibular nerve block upon speech. First, the effect was found to be subtle, limited, and manifest only in rapid, connected speech. In an array of utterances heavily weighted with consonant clusters, deterioration of articulation was noted by listeners in surprisingly little of the data. Transcribers agreed upon distortions in only 20% of the data (17% if the unaffected speaker is included), agreeing that there were no perceptible distortions in 67% of the utterances (71% when the unaffected speaker is included).

The effect was discovered to be limited to certain phonemes. It would be distorting the facts to report that the effect was largely with the fricatives, because although /s/ was by far the most common phoneme to deteriorate, and /z/ and /ʃ/ also underwent changes, there was seemingly no effect upon /ʒ/ or /θ/ and very little upon /f/ or /v/. The affricates were affected, but the plosives suffered very little, with only /t/ noticeably slurred. The nasals were not noticeably affected. There was some distortion of /r/ and /l/, but the most conspicuous effect remained the /s/.

Finally, the effect was found to be highly variable across subjects, a finding not mentioned in previous reports. Although all subjects reported complete loss of sensation in the anterior two-thirds of the tongue, the effects on speech ranged from completely unaffected to markedly affected. As one can see from referring to Table 3, the subjects varied from no effect to distortions in 55% of the utterances sampled, with significant variation among the subjects between the two extremes.

The high degree of intelligibility of all of the speakers in this investigation gives some weight to the theory that skilled speech may be largely under open loop control. At least it can be concluded that with only one sensory channel inhibited in its function, the motor sequencing of speech remains essentially intact. Skilled speech remains highly intelligible whether under conditions of auditory masking of one's own speech or conditions of oral sensory nerve block.

It may be that the systems used to monitor our own speech, specifically audition by air and bone conduction, vision, proprioception, and taction, are of primary importance during the learning of speech. After the process of speaking becomes relatively automatic and facile, we may monitor less and switch from one channel to another as we monitor.

It is unclear why there is such variability of nerve-block effect among subjects. It might be a difference in muscle use. It might reflect differences in the nerve block in sensory versus motor effects. Another possibility is individual variation in dependence upon sensory or auditory feedback. Some speakers may have developed a more open loop speaking system than other speakers.

An inherent variable in these investigations is the nerve-block injection itself. Despite subjective reports of loss of sensation, there are probable variations in depth of anesthesia and in the specific nerves affected. It would be advisable in future investigations to use more sophisticated methods of testing loss of sensation, both taction and kinesthesis. Electromyography should be used, in addition, to check motor function. The extent of the variability of speech effect after nerve-block anesthesia should caution researchers to avoid generalizations about the deterioration of articulation with sensory deprivation in the oral area.

REFERENCES

Borden, G. J. (1971) Some effects of oral anesthesia on speech: A perceptual and electromyographic analysis. Unpublished doctoral dissertation, City University of New York.

Chase, R. A. (1967) Abnormalities in motor control secondary to congenital sensory defects. In Symposium on Oral Sensation and Perception, ed. by J. F. Bosma. (Springfield, Ill.: Charles C Thomas).

Cook-Waite Labs, Inc. (1971) Manual of Local Anesthesia in General Dentistry, rev. 2nd ed. (New York).

Gammon, S. A., P. J. Smith, P. G. Daniloff, and C. W. Kim. (1971) Articulation and stress/juncture production under oral anesthetization and masking. J. Speech Hearing Res. 14, 271-282.

Locke, J. L. (1968) A methodological consideration in kinesthetic feedback research. J. Speech Hearing Res. 11, 668-669.

McCroskey, R. L. (1958) The relative contribution of auditory and tactile cues to certain aspects of speech. Southern Speech J. 24, 84-90.

McCroskey, R. L., N. W. Corley, and G. Jackson. (1959) Some effects of disrupted tactile cues upon the production of consonants. Southern Speech J. 25, 55-60.

McDonald, E. T. and L. F. Aungst. (1970) Apparent independence of oral sensory functions and articulatory proficiency. In Second Symposium on Oral Sensation and Perception, ed. by J. F. Bosma. (Springfield, Ill.: Charles C Thomas).

Ringel, R. L. and M. C. Steer. (1963) Some effects of tactile and auditory alterations on speech output. J. Speech Hearing Res. 6, 369-378.

Schliesser, H. F. and R. O. Coleman. (1968) Effectiveness of certain procedures of alteration of auditory and oral tactile sensation for speech. Percept. Motor Skills 26, 275-281.

Scott, C. N. (1970) A phonetic analysis of the effects of oral sensory deprivation. Unpublished doctoral dissertation, Purdue University.

Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia

Gloria Jones Borden,[+] Katherine S. Harris,[++] and Lorne Catena[+++]

Electromyographic recordings were made from the lip, tongue, and certain suprahyoid muscles of four normal adult speakers under normal conditions and under conditions of trigeminal nerve-block anesthesia. The mylohyoid muscle and the anterior digastric muscles which are innervated by motor fibers from the blocked nerve were usually depressed or inactive during the nerve-block condition. The assumption that the effects of this traditionally used nerve block are purely sensory seems unfounded. Other muscles are either depressed in activity during the block or more active than normal during the block. The amplitude of EMG recording depends upon depth and symmetry of anesthesia and upon the idiosyncratic reaction of the subject. Changes in muscle activity during the nerve block extend even to those muscles whose sensory and motor innervations cannot be affected by the block. Therefore, the effects observed indicate a more central effect or some compensatory reorganization.

[+]Haskins Laboratories, New Haven, Conn., and City College of the City University of New York.

[++]Haskins Laboratories, New Haven, Conn., and the Graduate Division of the City University of New York.

[+++]University of Connecticut Health Center, Farmington, Conn. Currently, Southern Illinois University, Edwardsville.

Acknowledgment: Part of this article summarizes a portion of a doctoral dissertation by the first author completed at the Graduate Division of the City University of New York under the direction of Katherine S. Harris (1971). The authors gratefully acknowledge the assistance of Victor Caronia, D.D.S., of Columbia School of Dental and Oral Surgery and Fredericka Bell-Berti of Montclair State College. Dr. Robert Ringel of Purdue University administered the sensory tests in Experiment II and offered many helpful suggestions throughout the work. Indispensable to these studies were Dr. Masayuki Sawashima and Dr. Hajime Hirose of the Faculty of Medicine, University of Tokyo, who inserted the electrodes. This research was supported in part by a grant from the National Institute of Dental Research to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

A series of studies during the 1950s and '60s dealt with the subject of the role of tactile feedback in speech. It was found that bilateral mandibular and infraorbital injections of anesthesia increased the number of judged errors in articulation of adult speakers (McCroskey, 1958; Ringel and Steer, 1963). The speech distortions were found to be subtle and were most evident in the production of fricatives and affricates (Scott, 1970; Borden, 1971; Gammon, Smith, Daniloff, and Kim, 1971). It was assumed by the investigators that the speech effect was primarily due to decreased sensory feedback as a result of blocking oral sensation from the tongue via the lingual nerve. A phonetic analysis of the speech effect under anesthesia revealed two factors which prompted further investigation; first was the variability of effect among speakers, with some subjects unaffected by the nerve block, although oral sensation was reported to be lost, and the second factor was the predominance of articulatory distortions among the sibilants and affricates, especially /s/ in consonant clusters, in those subjects who were affected (Borden, 1971). It was decided to study electromyographically the contraction of some of the muscles thought to be implicated in lingual movement under conditions of nerve block and under normal conditions.

Four separate electromyographic (EMG) experiments were conducted in an attempt to find out what happens to certain suprahyoid and tongue muscles as subjects speak under conditions of trigeminal nerve block.[1]

FIRST ELECTROMYOGRAPHIC STUDY

Since the nerve block seemed to produce an /s/ effect, muscles which are thought to contribute to tongue elevation were reviewed (Van Riper and Irwin, 1958; Hirano and Smith, 1967; Zemlin, 1968). The muscles which were accessible, clearly identifiable, and of interest for this study were the genioglossus, geniohyoid, mylohyoid, and the anterior belly of the digastric muscles. The orbicularis oris was included as a reference (Figure 1).

Method

The monopolar electrodes used were DISA concentric needle electrodes with a diameter of .45 mm. Needle placement was made through the cutaneous tissue under the chin to the depth required. Correct placement was checked by observation of an oscilloscope while protruding the tongue for genioglossus activity, saying "ta" for geniohyoid activity, lowering the mandible for digastric activity, and saying "ka" for mylohyoid activity. Correct placement was checked periodically throughout each run.

The subject for the first experiment was a normal adult speaker. There were two experimental conditions--without nerve block, and with bilateral

---

[1] These studies were conducted over a substantial period of time, during which electrodes, insertion techniques, and data analysis were substantially altered. In particular, the first experiment, in 1969, was performed under circumstances which permitted only relatively gross statements to be made about the results. Insertion techniques for the intrinsic tongue muscles were developed just before the third experiment was performed.

Figure I

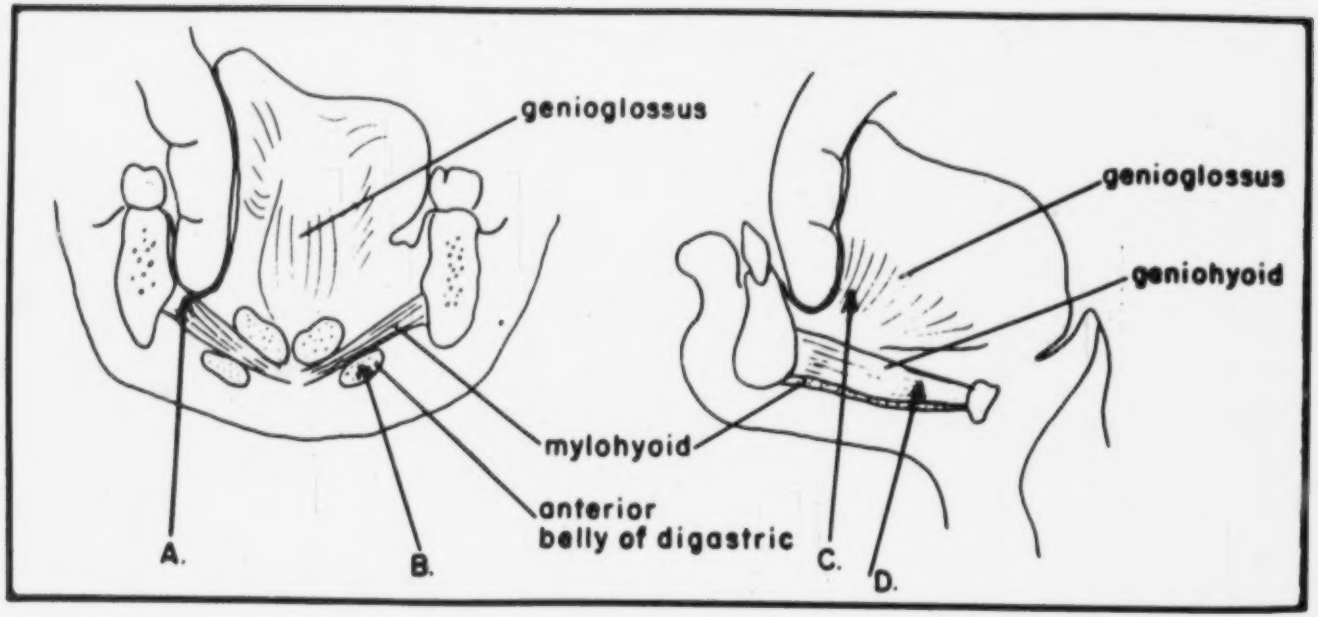


Figure 1: Muscles examined in first EMG study: front and sagittal views. Arrows indicate direction of needle insertions.

mandibular blocks. A total of 7.5 cc of 2% Xylocaine was injected by a dentist, 3 cc in each side and an additional 1.5 cc on one side. The technique was similar to that used by McCroskey (1958), the model for all previous studies. A partial run was recorded with a medial nasopalatine block of 1 cc and an anterior palatine block of 2 cc added, but this part of the study was not analyzed, as the speech effects were not noticeably different from the run with the bilateral mandibular blocks alone. It seemed that loss of sensation from the anterior portion of the hard palate and the alveolar ridge adds very little to the speech effect from the mandibular block.

For the EMG studies, material was selected from the utterances used in our previous work (Borden, Harris, and Oliver, 1973). Eleven utterances in sentence form, using the format "It could be the _____." were used to permit normally paced connected speech. Each utterance was represented twice in a randomized list of 22 utterances. There were 10 such lists, each individually randomized. Each utterance was spoken 20 times during the course of one run. The utterances were:

It could be the snowballs splashing.

It could be the cat's whiskers.

It could be the fixed sweater.

It could be the school blocks.

It could be the thirsty wasp.

It could be the sleeping taxi.

It could be the spider string.

It could be the squirrel nest.

It could be the rooster scratch.

It could be the spring grapes.

It could be the stove smell.

The 220 utterances for each run were printed and mounted on large cards which were flipped as the subject read them, with equal stress attempted on each of the final two words.

A 16-channel magnetic tape was produced, recording the electrical output of the muscles. Recordings were monopolar; that is, the voltage difference was recorded between the active tissue of the muscles and the inactive tissue of the earlobe. Some channels were used for audio signals, such as the utterances produced by the subject and the experimenters' comments for record-keeping. Each utterance was numbered by a pulse code laid down on the tape and eventually used for computer synchronization.

A visual record of the EMG and audio channels was made for locating and inspecting the individual tokens. Each utterance was represented 20 times during each run, and a single point in time, the line-up point, was selected so that all of the tokens of a single type could be averaged by computer for each electrode. The line-up point was chosen at a point of particular interest and marked on the simultaneous recording of the subject's audio recording.

Each tape was checked with five computer programs: to verify that the code pulses were in order, to set the gains of the playback amplifiers at levels appropriate for the analog-to-digital converter, to make control tapes of the line-up points and distances from point zero for each utterance, to set each EMG channel at the optimum level, and finally to average the data on the control tapes. The three runs were hand-plotted (Harris, 1970).

Results and Discussion

Inspection of the data revealed that, except for two muscles, the muscular activity recorded during speech under nerve-block conditions was similar in amplitude to that recorded during the normal condition. However, the activity observed on the oscilloscope of the mylohyoid muscle and the anterior belly of the digastric muscle after nerve-block injections dropped dramatically. The electrodes were checked and found to be in place, but as long as the anesthesia was effective those muscles were, in effect, paralyzed. The speech of the subject under nerve block revealed the typical mandibular block effect of distorted sibilants, the /s/ clusters being most prominently affected. For example, for the production of the utterance "sleeping taxi," Figure 2 shows the activity of the mylohyoid muscle and the anterior belly of the digastric during normal and nerve-block conditions. Graphs of all 11 utterances demonstrate the same drop in activity of these two muscles.

A closer look at the anatomy of the injection area suggests a reason for this effect. The mandibular injection which has traditionally been used for these studies deposits half of the solution in the area of the lingual nerve, then moves on to deposit the rest of the solution in the area of the inferior alveolar nerve. Just before the inferior alveolar nerve enters the mandibular foramen into the mandibular canal, it gives off the nerve fibers of what is known as the mylohyoid nerve, the only purely motor component of the otherwise sensory inferior alveolar branch of the trigeminal nerve. The mylohyoid nerve is motor to the mylohyoid muscle and to the anterior belly of the digastric muscle, the two muscles which dropped in activity during the nerve-block condition. The anatomy of the area is indicated in Figure 1, Part I.

The question was whether the inactivity of either of these muscles could have contributed to the noted speech deterioration. If the speech effect is primarily due to sensory loss, then loss of feedback from the tongue-tip region would probably be responsible. If it is due to motor loss, however, then the inactivity of the anterior belly of the digastric muscle and the mylohyoid muscle would probably be responsible.

The normal function of the anterior belly of the digastric muscle is to open the jaw. EMG data on this muscle, obtained by recording muscle activity during simple consonant-vowel-consonant (CVC) utterances, showed no action for /i/ and /u/ and a large peak for /ɑ/ (Harris, 1971). Since there was no perceptible speech effect of the nerve block upon vowels, and since the action of the anterior belly would not reasonably be expected to affect the apical gestures which deteriorated under the nerve block, it seems unlikely that its motor loss could have caused the speech effects observed.

The normal function of the mylohyoid muscle was found by both Harris (1971) and Smith (1970) to be highest for the production of /k/. Its contraction seems

Figure 2

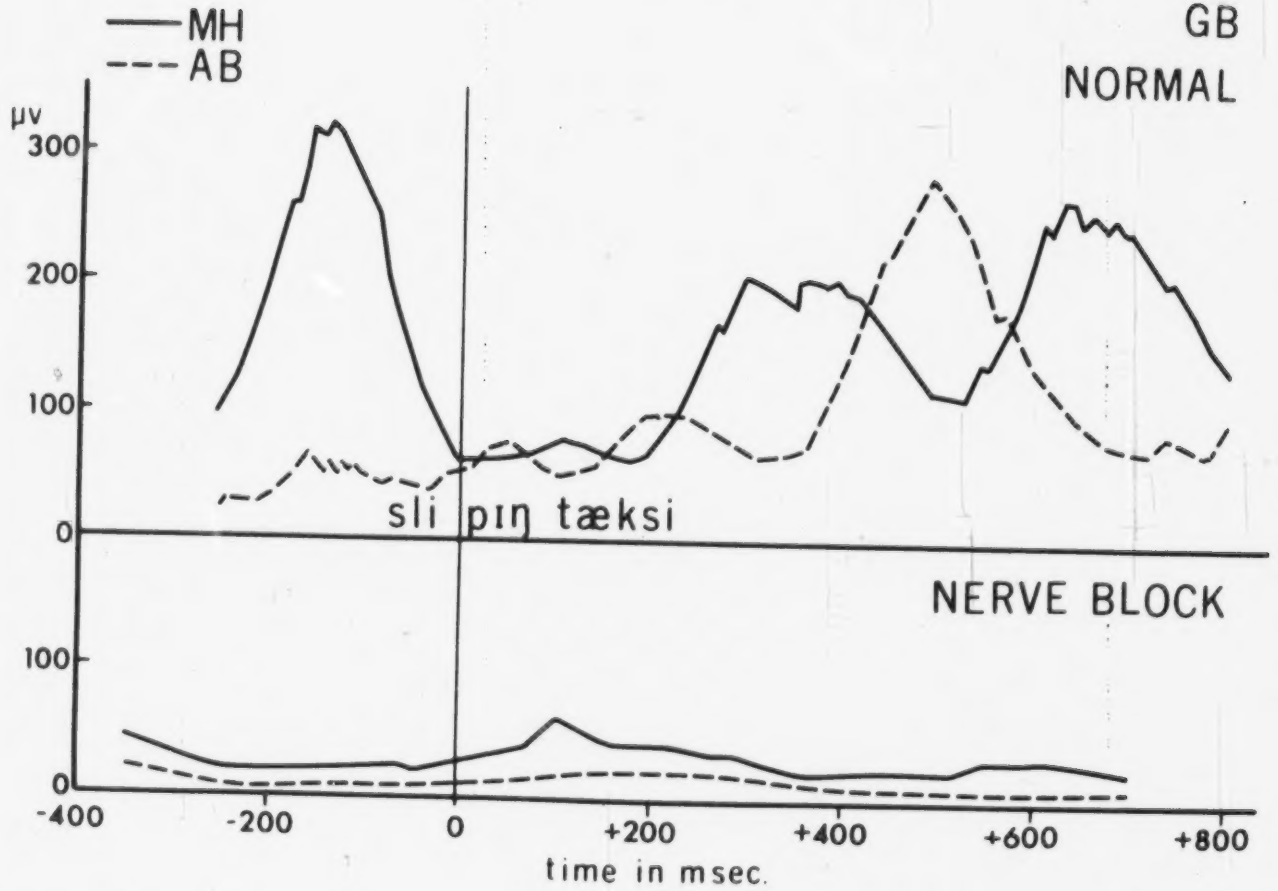


Figure 2: EMG recording of the mylohyoid muscle (MH) and the anterior belly of the digastric muscle (AB) during normal and nerve-block conditions, Experiment I.

to lift the body of the tongue. In the more complex utterances of the present study, it can be seen that the mylohyoid muscle peaked normally in preparation both for the /s/ consonant clusters and for the velars (Figure 3). Notice the activity at the beginning of "spring," "spider," and "string," and at the end of "grapes" and "string," in the normal condition. The drop in activity of the mylohyoid muscle during the nerve-block condition is obvious. The peaks of activity under normal speaking conditions, then, coincided with production of the segments that were distorted under the nerve-block condition, with the exception of the velars.

The velars were not perceived as distorted in the nerve-block condition. The production of /k/ remained intact, as had been reported in all previous nerve-block experiments. The explanation may lie in the comparatively gross production of /k/ and the fact that listeners accept for /k/ a less precise gesture than for /s/.

It seemed, therefore, that the effective paralysis of the mylohyoid muscle might reasonably be related to the speech effect, since, for this subject, the mylohyoid muscle appears to be important in lifting and steadying the body of the tongue for consonant clusters, especially those with /s/ (Table 1). This subject produces /s/ with the tongue tip down, making it imperative that the body of the tongue be raised to produce the friction. Deprived of motor ability in the mylohyoid and deprived of lingual sensation, the /s/ clusters were distorted. It is impossible to conclude which of these factors, if not both, is responsible for the distorted speech.

In summary, the clear conclusion of this first EMG experiment was that a motor component seemed to exist in what was previously assumed to be a sensory deprivation. The motor loss was evident in two of the suprahyoid muscles, the mylohyoid muscle and the anterior belly of the digastric muscle. One of these muscles, the mylohyoid, is normally active for this subject for /s/ clusters and velars. Since this subject produced /s/ with a high dorsum, it is reasonable to assume that the motor loss in the mylohyoid muscle may have contributed to the speech deterioration during anesthesia. However, the lack of effect on /k/ could not be unequivocally explained.

SECOND ELECTROMYOGRAPHIC STUDY

The purpose of the second EMG study was to verify the result of the first study that mylohyoid motor loss accompanies the distorted speech during the nerve-block condition, and also to study further the changes in muscle activity by comparing the muscle activity in normal speech with the potentials generated during nerve block.

There were two differences in procedure from the first study; first, since we wanted to find out if the motor loss was inevitable under the normal administration conditions of the bilateral mandibular block, the administrator (Dr. Catena) tried to avoid heavy infiltration of the mylohyoid nerve, within the bounds of the previously described injection technique. Second, we wanted to look at block conditions which more nearly corresponded to those used by Scott (1970).

Figure 3

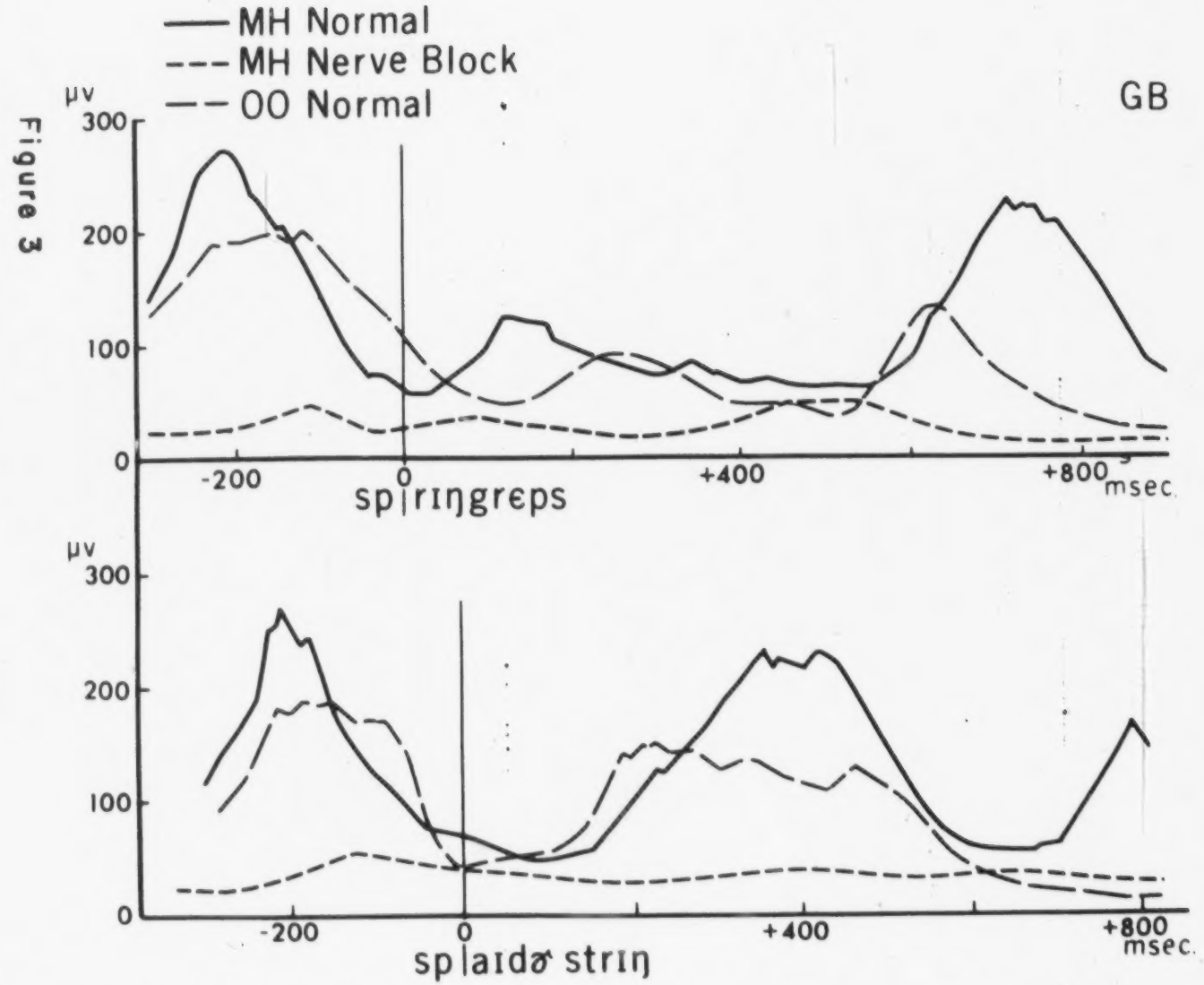


Figure 3: Mylohyoid muscle (MH) peaks under normal conditions for /s/ consonant clusters and for velars. The orbicularis oris (OO) is included as a reference, Experiment I.

TABLE 1: Peak values in microvolts for mylohyoid muscle in first EMG experiment during nerve-block and normal conditions.

	spring	rapes	
Normal	345	155	285
NB	30	35	20
msec	(-225)	(125)	(715)

	rooster	scratch		
Normal	175	200	310	370
NB	30	40	40	20
msec	(-775)	(-440)	(-125)	(325)

	cats	whis	kers	
Normal	315	355	380	370
NB	35	40	40	20
msec	(-800)	(-505)	(-140)	(200)

	fixed	sweater	
Normal	485	210	140
NB	45	45	15
msec	(45)	(325)	(585)

	thirsty	wasp	
Normal	185	310	
NB	30	35	
msec	(-855)	(-255)	

	school	blocks	
Normal	380	400	
NB	50	30	
msec	(-145)	(640)	

	stove	smell	
Normal	335	355	
NB	30	50	
msec	(-215)	(325)	

	squirrel	nest	
Normal	215	150	
NB	50	25	
msec	(-175)	(635)	

	snowballs	splashing	
Normal	415	340	430
NB			
msec	(-140)	(500)	(900)

(/ng/ not plotted)

	spider	string	
Normal	355	300	210
NB	35	40	25
msec	(-210)	(365)	(790)

	sleeping	taxi	
Normal	425	265	355
NB	30	40	40
msec	(-155)	(300)	(635)

Method

It was necessary for technical reasons to use a second subject for this experiment. The material consisted of 30 utterances in the frame "the _____." They were randomized into four lists repeated alternatively four times, making 16 lists of 30 utterances each. Fifteen of the utterances were chosen from the Scott (1970) list in an attempt to observe the muscle changes in the distorted speech which might explain the phonetic changes that she had described. The other 15 utterances were words selected from the sentences in the first study and from the perceptual study. Two runs were produced, the first under normal conditions, the second under blocked condition.

The electrodes were 0.002-in wires hooked to remain in place. Correct placement was checked by observing the oscilloscope while lifting the tongue for genioglossus activity, tensing the floor of the mouth while relaxing the tongue for geniohyoid activity, saying "ka" for mylohyoid activity, opening the mouth with jaw effort for anterior belly of digastric activity, saying "pa" for orbicularis oris activity, and lifting the head or opening the mouth under pressure for sternohyoid activity. The genioglossus and geniohyoid were also checked during swallowing, as their activity differs in timing (Hirose, 1971). Electrodes were placed in both sides of the mylohyoid muscle and in both anterior bellies of the digastric muscle.

After the normal run, a total of 7.5 cc of 2% Xylocaine was injected into the oral region of the subject. A summary of the injections is given in Table 2. Details on the technique used may be found in a standard reference of dental anesthesia (e.g., Cook-Waite Labs, 1971).

TABLE 2: Injections of anesthesia administered in the second EMG study.

Cranial Nerve	Branch	Amount of Solution	Location of Injection	Area of Sensation
V (mand.)	Inf. Alveol. n. Lingual n.	1.5 cc ea. side	pterygomand. triangle	mand. alv. ridge, lip, gum ant. 2/3 tongue
V (mand.)	Long Buccal n.	.5 cc ea. side	1st molar	buccal
V (max.)	Infraorbital Ant. Sup. Alv. Middle Sup. Alv.	.5 cc ea. side	infraorbital foramen	upper lip alv. ridge ant. teeth
V (max.)	Nasopalatine n.	.5 cc midline	post. to central incisors	ant. 1/3 palate
V (max.)	Post. Sup. Alv. n.	.5 cc ea. side	palate 3rd mol.	post. 2/3 palate

A rough check of two-point discrimination was made, and when the experimenters and subject were satisfied that sensation was lost in the tongue and the palate, Ringel's (Ringel, House, Burk, Dolinsky, and Scott, 1970) 55-item oral discrimination test of 10 plastic forms was administered. When the subject had returned to normal, the Ringel test was again administered. The subject made nine errors in normal condition and fifteen errors in the nerve-block condition, the difference being errors of shape, not size. Confusion of shape occurred three times in normal condition and nine times in nerve-block condition. Nevertheless, the experimenters were surprised that there was so little difference in performance on this test. It was noted that the subject used the usual tongue manipulations during normal condition but relied on deep pressure against the palate when sensation was decreased. This technique was reported as the method used by successful subjects in the study on the effect of anesthesia on oral stereognosis (Mason, 1967).

The multichannel magnetic tapes produced for each of these runs were analyzed in much the same way as the first experiment. There were some refinements in the computer programs. A concise description of the analysis procedure is reported by Port (1971).

Results and Discussion

The most conspicuous result of the second EMG experiment was that the subject's articulation remained clear during the condition of nerve block. The speech sounded as acceptable under the nerve-block condition as under the normal condition. The utterances were louder under nerve block and produced with what might be described as over-articulation.

This variability of nerve-block effect among subjects was observed during the perceptual part of this series of studies. It is unclear why there was no speech effect. It might be a difference in muscle use, as this subject produces /s/ with tip of the tongue raised and might not rely on mylohyoid muscle activity as much as the first subject, who produces /s/ with the dorsum of the tongue raised, keeping the tip down. Another explanation for the lack of speech effect might be a difference in anesthesia, either in amount or in technique of injection.

Following the first EMG experiment, the investigators were particularly interested in this second study in the activity of the mylohyoid muscle. Since there were bilateral placements of electrodes in both the mylohyoid muscle and the anterior belly of the digastric muscles, the investigators had an opportunity to study the activity on both sides of these muscles. During the normal run, before the injections of anesthesia, the mylohyoid and the anterior belly showed activity similar to the first subject. The anterior belly peaked for mouth opening and the mylohyoid for velar gestures and somewhat for the /s/ clusters.

During the condition of nerve block, however, there was a change in activity in both muscles on the right side. The right anterior belly of the digastric was in all cases less active than normal after anesthesia. The right mylohyoid was consistently less active than normal for velar gestures, but for the /s/ clusters, it was sometimes less active and sometimes more active than normal. The gain in activity for the nonvelar gestures offset the loss for /k/. The decreased activity on the right side in this experiment was not as pronounced as it had been in the first EMG study, indicating that the attempt on the part of

the dentist to avoid the motor mylohyoid nerve was partially successful. The limited effect on the right side was presumed by the investigators to be the result of some infiltration of the anesthetic in the area of the mylohyoid nerve, especially affecting nerve fibers which are motor to the digastric muscle.

In contrast with the instances of decreased activity observed on the right side of the mylohyoid and anterior belly of the digastric muscles, the left side of these muscles was usually more active than normal while the anesthesia was in effect. Figure 4 demonstrates the asymmetry of effect. The right peak in each of the four graphs represents the labial closing for /p/ in "duckpond." It can be seen that the right side of both muscles was quite active during normal speech but dropped in activity during speech with nerve block. Figure 4 also shows that by contrast both muscles on the left side were more active than normal under block.

Figure 5 summarizes the activity for each muscle during the nerve-block condition relative to its normal activity. Taking the normal peak amplitude in microvolts of each electrode during the central 400 msec around each line-up point as 100%, the percentage of normal activity was computed for the peak amplitude during the nerve-block condition for each utterance. The average of the 30 percentages is represented in the figure.

In this figure, muscles have been arranged into three groups--those where the bilateral mandibular block might be expected to have a motor effect, those where the effects of the bilateral mandibular block should be sensory, and those where the effects observed lie outside the field of the bilateral mandibular block, though within the field of the other blocks performed. This arrangement is intended for comparison with the results of Experiments 3 and 4, although it is not logical for the data of this experiment. The meaning of the term "sensory" is discussed at greater length in the final section.

Overall, there is a widespread change in the activity of the muscles sampled, both those which are directly associated with tongue movement and those which are not. It seemed desirable to try to separate motor and sensory effects of the mandibular blocks.

THIRD ELECTROMYOGRAPHIC STUDY

Since the first two EMG studies with nerve block appeared to reflect the results of a mixed nerve block, that is, the injection of anesthesia apparently affected both motor and sensory fibers of the trigeminal nerve, it was considered important to make further attempts to separate these factors. The third and fourth investigations were designed to anesthetize the lingual nerve alone, producing a purely sensory block, and to anesthetize the mylohyoid nerve alone, producing a purely motor block. The third EMG study was an attempt to investigate the effects of the lingual nerve block.

Furthermore, since we found that there was a change in EMG output under the nerve-block condition even without a speech effect, we wanted to be sure that we used only subjects whose speech was perceptually distorted under the block condition.

Figure 4

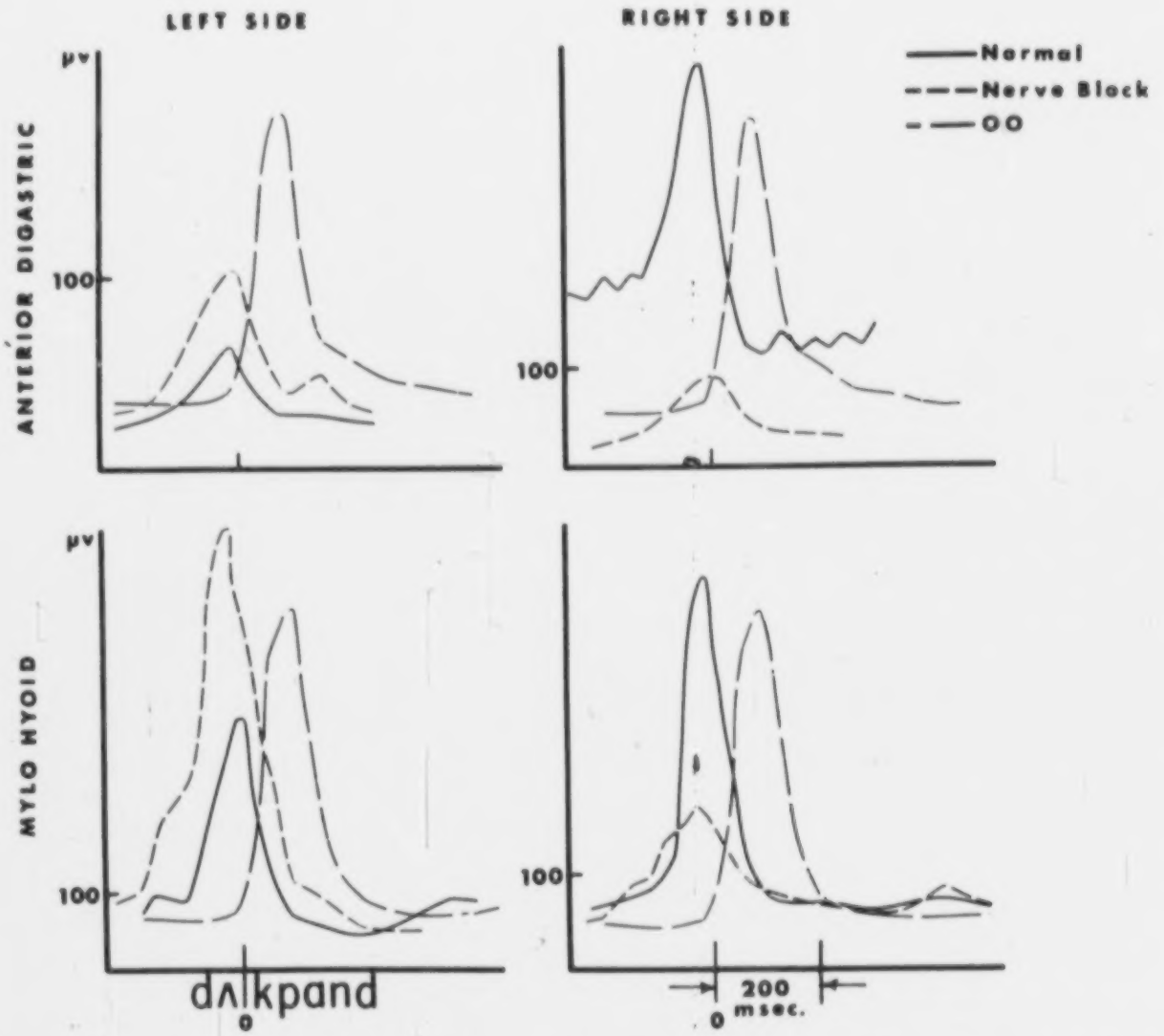


Figure 4: Decreased right side activity and increased left side activity during nerve block of the mylohyoid and anterior belly of the digastric muscles. The orbicularis oris (OO) is included as a reference, Experiment II.

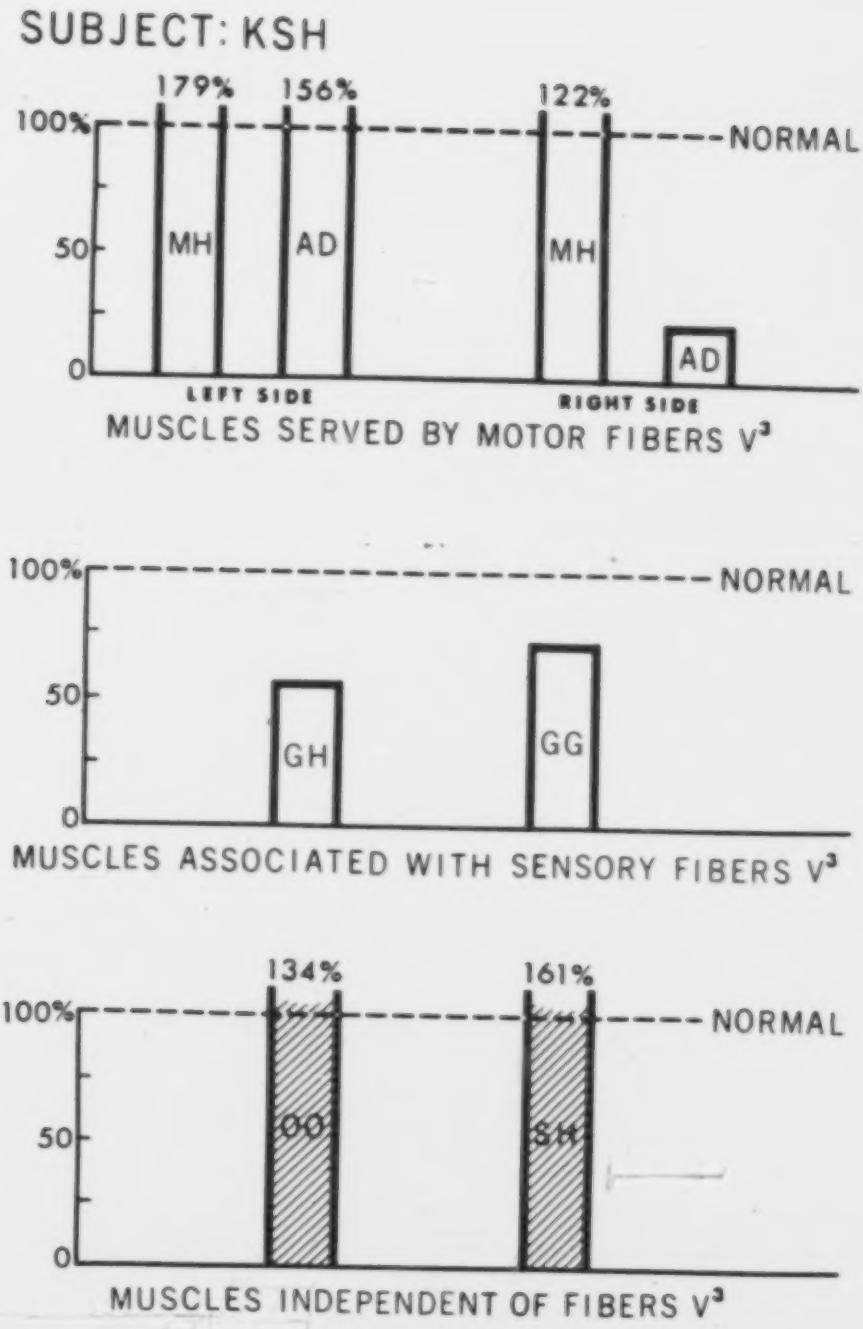


Figure 5: Mean percentages of normal peak EMG amplitudes in microvolts for muscles during nerve block, Experiment II.

Method

The 11 sentences used in the first EMG study were repeated 18 times each in nine randomized lists by two subjects. Stress was placed on the first key word, "It could be the sleeping taxi." Subjects were selected by conducting a short trial run during which four candidates were given the routine bilateral mandibular injection of 2% Xylocaine with 1:100,000 epinephrine. Of the three candidates who evidenced speech distortions during the nerve block, two, both male speakers of English, were chosen as subjects. Tests of two-point discrimination of the tongue using a Downes aesthesiometer and of oral stereognosis using the National Institute of Dental Research forms were conducted during normal and blocked conditions. By a slight modification of the injection technique an attempt was made to block only the lingual nerve using 1 cc Xylocaine with 1:100,000 parts epinephrine on each side. During the normal condition subject DL could make accurate two-point discriminations at 3 mm in most cases, requiring up to 4 mm separation in some instances at the anterior part of the tongue and up to 1 cm separation at some points on the posterior part of the tongue. During the nerve-block condition, however, DL failed to discriminate accurately in five out of eight two-point placements even when point separation reached 1.5 cm. Oral stereognosis ability declined also. Eight errors out of 18 were scored during the normal condition and 14 errors were scored during the nerve-block condition.

The second subject PN made few errors of two-point discrimination at 3 mm normally but reported no sensation at all 16 placements during the blocked condition. Three errors of identification of the forms normally were increased to 13 errors out of 18 possible identifications during the nerve block. The investigators presumed success in lingual nerve isolation in the case of DL, as sensation was reported lost in the anterior two-thirds of the tongue but remained on the lower lip and gingivae. The effect upon subject PN was less clear, as there was some loss of sensation in the lower alveolar ridge and lower lip, indicating a partial block of the inferior alveolar nerve.

EMG recordings were made from the superior orbicularis oris, the anterior genioglossus, bilateral placements in the mylohyoid, and in the anterior belly of the digastric muscles. New in this experiment were electrode placements in the superior longitudinal muscle of the tongue. Bilateral insertions were made approximately 1 cm from the midline and 1 cm from the tip of the tongue. The insertions were superficial, with an estimated depth of 2 to 3 mm. The hooked wires were located about 1 cm posterior to the point of insertion.

The method of recording and analysis of data was the same as for the second EMG study.

Results and Discussion

Again, results may be described by grouping the muscles investigated according to whether the block effects on them may be considered to be sensory, motor, or indirect. The results indicate first, that the nerve block produced a rather dramatic effect on the contraction of the intrinsic tongue muscles from which we recorded. Subject DL evidenced a drop in activity during the nerve-block condition. The superior longitudinal muscle normally peaks for /θ/ and /l/. Both left and right electrode placement showed decreased activity, as did the genioglossus, another tongue muscle. Subject PN, however, reacted quite differently

to the nerve block. Superior longitudinal activity was depressed on the right side in a manner similar to the first subject, but the left electrode, in contrast, recorded much more electrical activity during the nerve-block condition than during the normal condition. The genioglossus muscle was also more active than normal. The effect of the nerve block in tongue muscles was generally depression of activity in subject DL; in subject PN, one side depressed and the other side evidenced greater effort under nerve block.

The nerve block also produces decided changes in EMG activity in muscles served not by sensory nerves involved in this nerve block but by motor nerves. The mylohyoid muscle, which normally contracts for /k/, showed greatly decreased activity on the left side. Subject PN showed almost total bilateral inactivity of this muscle for each token of each utterance type. Both subjects showed depressed anterior digastric activity during the nerve-block condition.

There is a change in the activity of a muscle whose innervation lies entirely outside the field of the block--the superior orbicularis oris. For subject DL it was somewhat depressed in amplitude during nerve block, but for subject PN it was much more active. Examples of orbicularis oris activity are shown in Figure 6. Changes in the level of activity peaks for /p/ can be seen in the block condition.

When the absolute peak values in microvolts during nerve block are compared to the normal peak values, and the percent of normal is averaged for each muscle, we can see the pooled difference from the normal condition which the nerve block produces. Again, only the peaks close to the averaging lineup point were chosen for analysis. Figure 7 shows that for subject DL, the nerve block produced a consistently depressed state of activity. The general depression extended even to muscle fibers that should have been completely unaffected by the block. Subject PN, however, has a far more complex pattern of activity over a wide range of muscles (Figure 8).

To summarize the effects of the nerve block in this experiment, the first class of muscles, those innervated by motor fibers from the blocked nerve, were consistently depressed or inactive. Thus, it seems that despite the attempt to anesthetize the lingual nerve alone, there is evidence of infiltration of the anesthesia. The next two classes of muscles, those presumably associated with sensory fibers from the blocked nerve and those which should be independent of the blocked nerve, were sometimes less active, sometimes more active, depending upon the side of electrode placement and upon the idiosyncratic reaction of the subject.

FOURTH ELECTROMYOGRAPHIC STUDY

If it is difficult to isolate the lingual nerve without affecting the motor fibers of the mylohyoid nerve, it is possible perhaps to anesthetize the mylohyoid nerve alone, producing a motor block while leaving the sensory fibers of the lingual nerve unaffected. This was the purpose of the final EMG study.

Method

The method was a repetition of the third study with the same electrode placements, the same utterance lists, and one of the same subjects, PN. The difference was that .5 cc of 2% Xylocaine with 1:100,000 parts epinephrine was

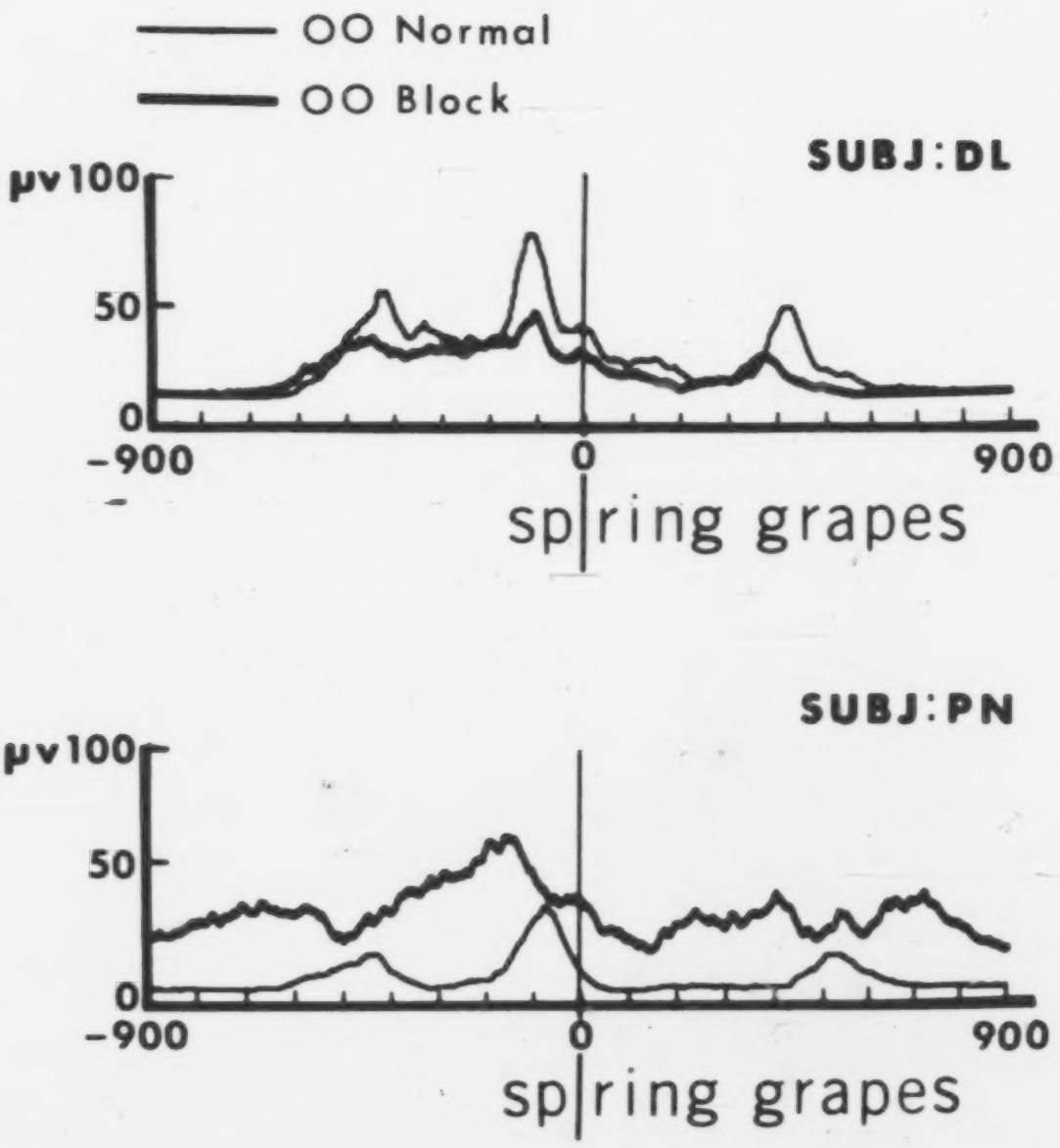


Figure 6: Reduced activity of the orbicularis oris during nerve block for one subject and increased activity for another subject, Experiment II.

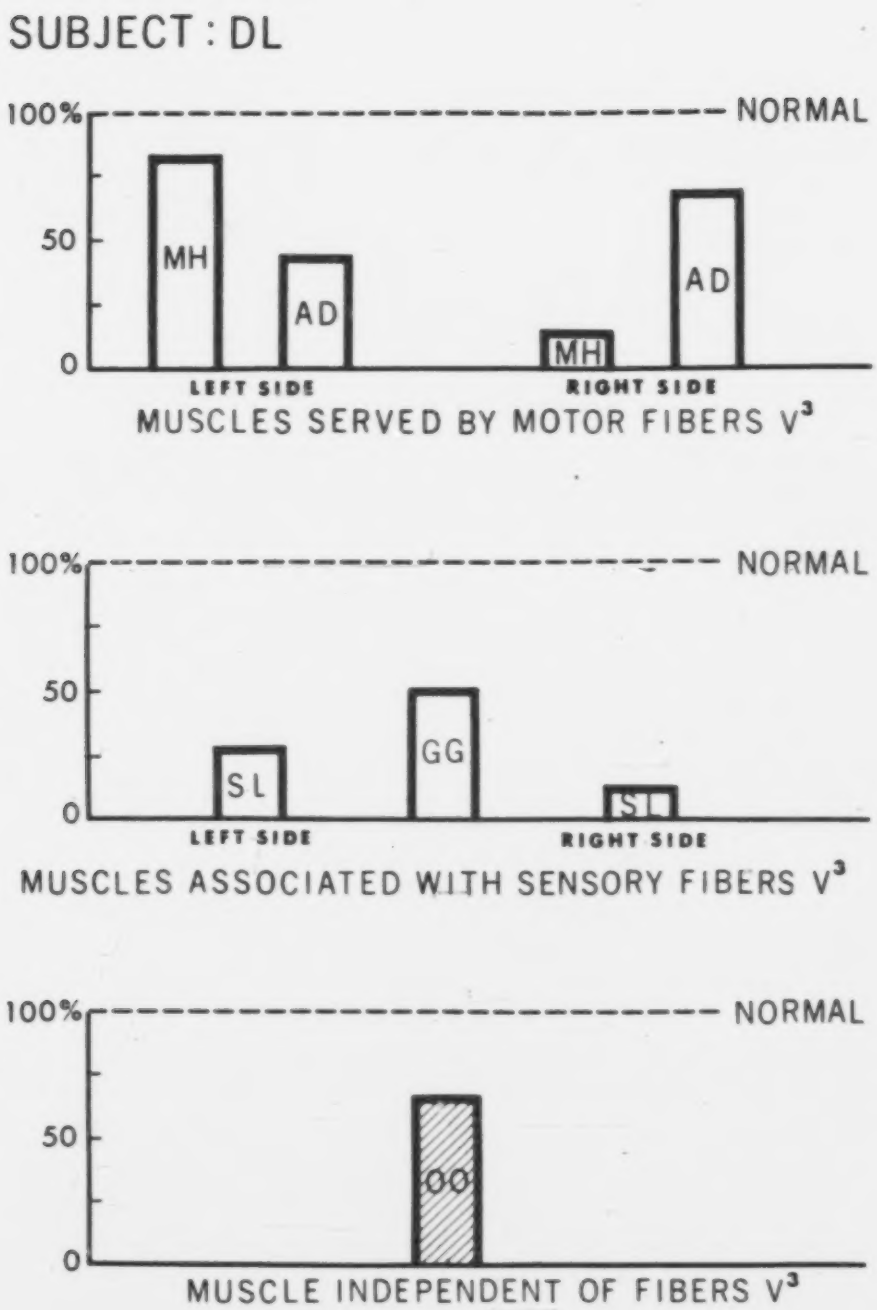


Figure 7: Mean percentages of normal peak EMG amplitudes in microvolts for muscles during nerve block, subject DL, Experiment III.

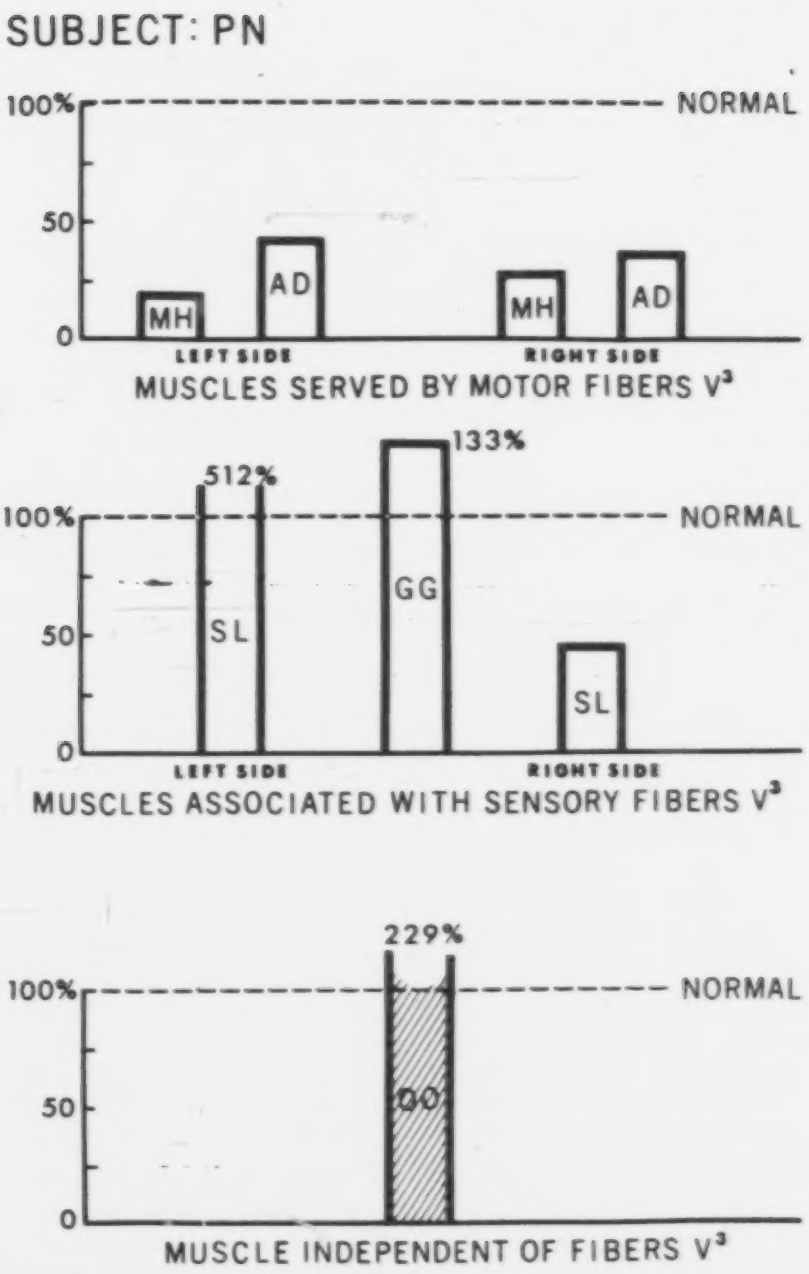


Figure 8: Mean percentages of normal peak EMG amplitudes in microvolts for muscles during nerve block, subject PN, Experiment III.

injected on each side at the juncture of the lingual mucosa and the floor of the mouth at the level of first molar. Data analysis was the same as for the second and third experiments.

Results and Discussion

The injection of the anesthesia directly into the mylohyoid muscle on each side produced much more of an effect on the left side, which was completely inactive after the block, than on the right side, which was depressed in activity but remained active (Figure 9). The left anterior digastric electrode was misplaced.

The intrinsic tongue muscles did not greatly alter activity, although the left side of the superior longitudinal was generally less active than normal and the right side more active than normal. The orbicularis oris was also somewhat more active during the nerve-block condition.

The subject's speech remained as well articulated as normal. The subject was not conscious of any sensory or motor changes as a result of the injection of anesthesia.

It seems to be as difficult to obtain a bilateral motor effect as it is to obtain a purely sensory nerve block. There were changes in the amplitude of the muscles sampled, however, even when there was little or no sensory loss.

SUMMARY OF THE EMG STUDIES AND DISCUSSION

Although the traditional bilateral mandibular nerve block often produces distortions in some of the gestures of rapid, connected speech, there is evidence that the effect may have both motor and sensory components. This was indicated by the total inactivity of the mylohyoid muscle and the anterior belly of the digastric muscle in the first study. The second study demonstrated the possibility of compensatory activity coupled with a lack of perceptible effect of the nerve block upon the articulation of speech. The third study confirmed the finding of the motor effect of the nerve block and increased the evidence of compensatory reorganization. Furthermore, the results demonstrated nerve-block effects upon muscles whose innervation is independent of the nerves involved. Increased activity under nerve block of muscles which are not served by either sensory or motor fibers of the anesthetized nerve indicates either a general reorganization of activity in an effort to compensate for some motor or sensory loss, or a more central effect of the anesthesia. It does not seem, from the results of the fourth study, that the motor effect alone is sufficient to distort the speech, although the fourth study also shows that there is some EMG reorganization without any evidence of the normal sensory effects of the block.

At this point, it seems worthwhile to try to reassess the results of these experiments in the light of the explanations usually offered for the nerve-block effect.

The primary reason for the effect may be motor, as we have previously suggested (Harris, 1970; Borden, 1972). On anatomical grounds, it is plausible that the block would affect motor innervation; indeed, it is quite difficult to make the sensory block while avoiding the motor innervation of two of the

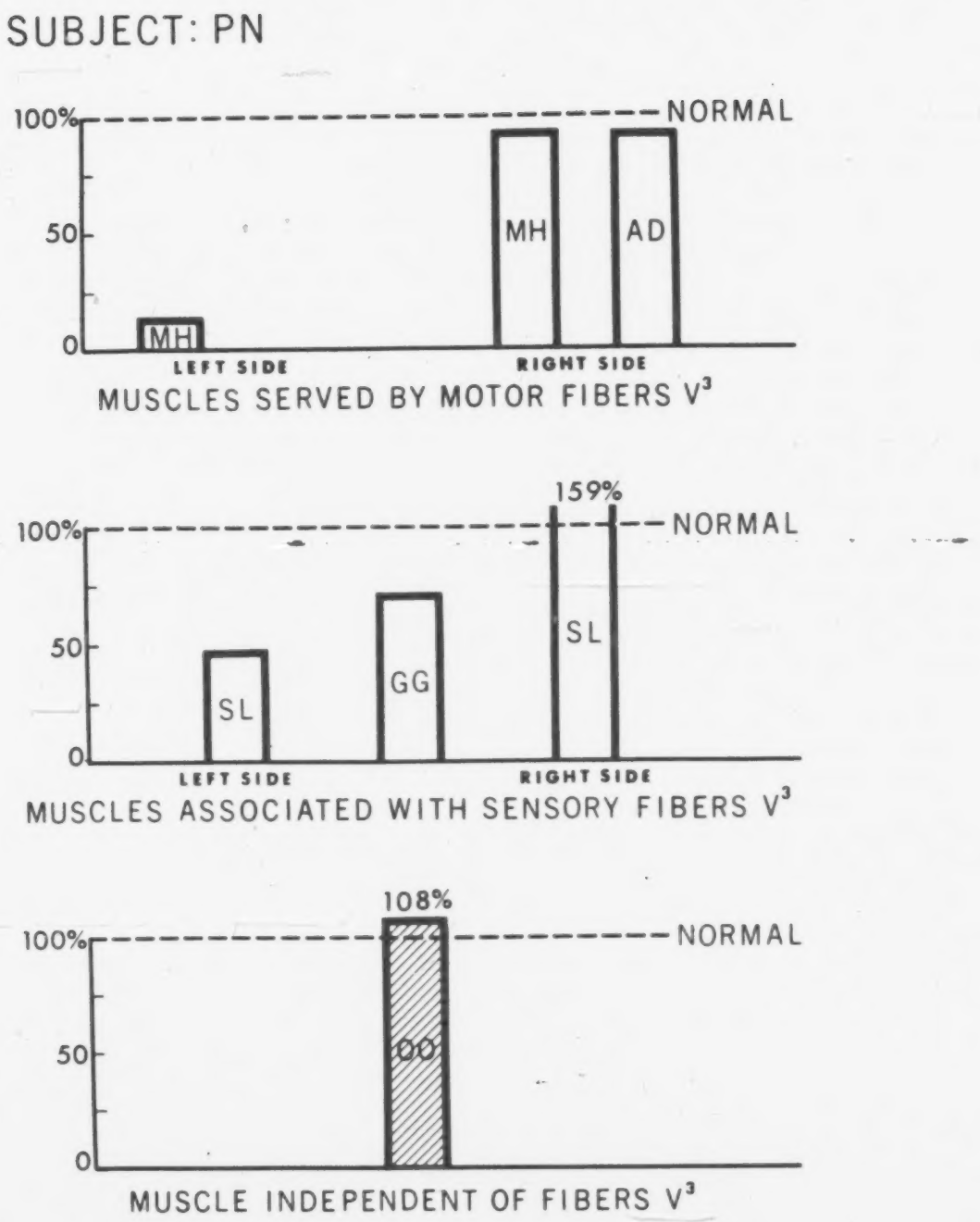


Figure 9: Mean percentages of normal peak EMG amplitudes in microvolts for muscles during nerve block, subject PN, Experiment IV.

muscles, the mylohyoid and the anterior belly of the digastric. However, the pattern of affected consonants makes a primary motor cause for the block effect unlikely. We would expect that inactivity of the mylohyoid muscle alone would make /k/ the most affected consonant; in fact there is general agreement that this consonant is spared. Furthermore, as we showed in Experiment IV, a block of the mylohyoid nerve will not apparently produce a perceptible speech distortion, at least in gross terms.

The most traditional explanation of the speech effect is that it is a consequence of decreasing sensory feedback from the oral area--either tactile or proprioceptive or both.

The "tactile" explanation is that a block of the lingual nerve cuts sensation from the surface of the tongue, which leads to imprecision in its placement. Again, the pattern of affected consonants makes the explanation somewhat implausible; in this case the consonants /t/, /d/, and /n/ should be maximally affected; they are not. Turning to the experiments reported above, the muscles most affected should be the superior longitudinal intrinsic muscles of the tongue, which lie closest to the numbed lingual surface. There is no evidence that their activity pattern is more, or less, affected than that of muscles lying deeper in the tongue body, or, indeed, of muscles lying outside the field of the block entirely. A simple tactile explanation does not seem tenable.

Another explanation for the block effect is that it causes interference with the proprioceptive return from muscle spindles in the tongue. If each muscle adjusts to a fixed length based on the return from its own stretch receptors, as has been described by MacNeilage (1970), then interference with this pathway should have serious effects on speech. Traditionally, it had been assumed that the lingual nerve carried proprioceptive as well as tactile information from the anterior two-thirds of the tongue, because the hypoglossal nerve has no sensory root (Blom, 1960). Studies in rhesus monkeys by Bowman and Combs (1968) would indicate that nerve fibers from muscle spindles in the tongue do course along the hypoglossal nerve for part of the way and then cross to join some cervical nerves. If this is the case in humans, the block spares proprioceptive feedback, since the injection site does not lie on the pathway of the hypoglossal nerve. If, on the other hand, proprioceptive feedback is carried in the lingual nerve, we would expect that the tongue muscles would be affected by the mandibular block, but not muscles outside the tongue, as we found in our third study.

Taking these results together, it would appear that any sensory effects of the block must be rather general. The system might be responding to an altered pattern of information sent back to the central nervous system with a changed motor output which affects muscles whose sensory feedback is normal--that is, there is not muscle-specific correction. These changes are most likely to alter those consonants which require the greatest degree of articulatory finesse.

Everyone writing on this effect recently has noted that the effect is restricted to a small class of consonants. The restricted results of all these studies provide us with some insight into the small size of the effect. The EMG signals may change size radically under the block; they do not seem to change their temporal relationship to each other. Changes in relative timing of the muscle gestures would produce far more devastating effects on articulation.

Recent work by Scott and Ringel (1971) has shown that the speech of subjects under block does not resemble that of a group of dysarthric speakers studied by Lehiste (1965) and Tikofsky and Tikofsky (1964). Their argument is that the effects cannot be motor, and hence, must be sensory.

Another possibility which probably should be considered is that the effect may be due to an additional factor, a generalized depression of central activity caused by the local anesthesia. Drowsiness is a well-known side effect of Xylocaine. Pharmaceutical studies indicate that local anesthetics may appear in considerable quantities in the blood stream (de Jong, 1968), and an effect upon speech is one clinical sign of a rising level of anesthetic in the blood. Furthermore, it has been shown that local anesthetics readily cross the blood-brain barrier (Usubiaga, Moya, Wikinsky, and Usubiaga, 1967). It is possible that a slight loss of central control may relate more directly to the slurring of speech than either the motor or sensory effects evidenced at the periphery. The speech effect, when it does exist, sounds perceptually very like 'drunk' speech. 'Drunk' speech is accepted as a consequence of the alcohol having crossed the blood-brain barrier to affect the central control of speech.

Whatever the cause of the nerve-block effect, it remains an important experimental technique because it is one of the few experimental means we have of altering speech production in normal adult speakers. Further work should be directed towards exploring the alternatives of general central effect versus sensory deprivation. Furthermore, EMG studies should be aimed at exploring other blocks to see if the patterns of their effects are similar to those of the bilateral mandibular block.

REFERENCES

Blom, S. (1960) Afferent influences on tongue muscle activity. Acta Physiol. Scand. 49, Suppl. 170, 1-97.

Borden, G. J. (1971) Some effects of oral anesthesia upon speech: A perceptual and electromyographic analysis. Unpublished Ph.D. dissertation, City University of New York.

Borden, G. J. (1972) Some effects of oral anesthesia upon speech: An electromyographic investigation. Haskins Laboratories Status Report on Speech Research SR-29/30, 27-47.

Borden, G. J., K. S. Harris, and W. Oliver. (1973) Oral feedback, part I: Variability of the effect of nerve-block anesthesia upon speech. Haskins Laboratories Status Report on Speech Research SR-34 (this issue).

Bowman, J. P. and C. M. Combs. (1968) The discharge patterns of lingual spindle afferent fibers in the hypoglossal nerve of the rhesus monkey. Exp. Neurol. 21, 105.

Cook-Waite Labs, Inc. (1971) Manual of Local Anesthesia in General Dentistry, rev. 2nd ed. (New York).

de Jong, R. H. (1968) Local anesthetic seizures. Surg. Dig. 3, 30.

Gammon, S. A., P. J. Smith, R. G. Daniloff, and C. W. Kim. (1971) Articulation and stress/juncture production under oral anesthetization and masking. J. Speech Hearing Res. 14, 271-282.

Harris, K. S. (1970) Physiological measures of speech movements: EMG and fiberoptic studies. ASHA Reports 5, 271-282.

Harris, K. S. (1971) Action of the extrinsic musculature in the control of tongue position: Preliminary report. Haskins Laboratories Status Report on Speech Research SR-25/26, 87-96.

Hirano, M. and T. Smith. (1967) Electromyographic study of tongue function in speech: A preliminary report. U.C.L.A. Working Papers in Phonetics 7, 46-56.

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and techniques. Haskins Laboratories Status Report on Speech Research SR-25/26, 73-86.

Lehiste, I. (1965) Some acoustic characteristics of dysarthric speech. Bibl. Phone. Fasc. 2.

MacNeilage, P. N. (1970) Motor control of serial ordering in speech. Psychol. Rev. 77 182-196.

Mason, R. M. (1967) Studies of oral perception involving subjects with alterations in anatomy and physiology. In Symposium on Oral Sensation and Perception, ed. by J. F. Bosma. (Springfield, Ill.: Charles C Thomas).

McCroskey, R. L. (1958) The relative contribution of auditory and tactile cues to certain aspects of speech. Southern Speech J. 24, 84-90.

Port, D. K. (1971) The EMG data system. Haskins Laboratories Status Report on Speech Research SR-25/26 67-72.

Ringel, R. L., A. S. House, K. W. Burk, J. P. Dolinsky, and C. M. Scott. (1970) Some relations between orosensory discrimination and articulatory aspects of speech production. J. Speech Hearing Dis. 35, 3-11.

Ringel, R. L. and M. D. Steer. (1963) Some effects of tactile and auditory alterations on speech output. J. Speech Hearing Res. 6, 369-378.

Scott, C. M. (1970) A phonetic analysis of the effects of oral sensory deprivation. Unpublished doctoral dissertation, Purdue University.

Scott, C. M. and R. L. Ringel. (1971) The effects of motor and sensory disruptions on speech: A description of articulation. J. Speech Hearing Res. 14, 819-828.

Smith, T. J. (1970) A phonetic analysis of the function of the extrinsic tongue muscles. Unpublished doctoral dissertation, University of California at Los Angeles.

Tikofsky, R. S. and R. Tikofsky. (1964) Intelligibility measures of dysarthric speech. J. Speech Hearing Res. 7, 325-333.

Usubiaga, J. E., F. Moya, J. A. Wikinsky, and L. E. Usubiaga. (1967) Relationship between the passage of local anesthetics across the blood-brain barrier and their effect on the central nervous system. Brit. J. Anaesth. 39, 943-946.

Van Riper, C. and J. V. Irwin. (1958) Voice and Articulation. (Englewood Cliffs, N. J.: Prentice-Hall).

Zemlin, W. R. (1968) Speech and Hearing Sciences: Anatomy and Physiology. (Englewood Cliffs, N. J.: Prentice-Hall).

Laryngeal Control in Korean Stop Production

Hajime Hirose,[+] Charles Y. Lee,[++] and Tatsujiro Ushijima[+++]

In Korean there is a three-way distinction in both manner and place of articulation that differentiates nine stop consonant phonemes. Linguists have disagreed about the manner classifications, describing them phonetically in various ways for initial position. Thus, Category I is characterized as voiceless, tense, long, strong, forced, and/or glottalized; Category II is voiceless, lax, slightly aspirated, and/or weak; Category III is voiceless, heavily aspirated, and lax according to some phoneticians but tense according to others (Martin, 1951; Umeda and Umeda, 1965; Abramson and Lisker, 1972). It is also known that the Category II stop typically becomes voiced intervocalically.

Much has been published in an effort to clarify the acoustical and physiological properties that differentiate these three manner categories. Among those, Lisker and Abramson (1964) made an acoustical investigation into various languages and showed that values of voice onset time (VOT), the temporal relation between stop release and onset of glottal pulsing, provide the most useful measure for differentiating various conditions of voicing and aspiration in word-initial stops. They noted, however, that Korean is peculiar in that the resolution of VOT values between Categories I and II is not clearcut but shows overlapping values, while Category III is well separated from the others. Similar observations have been reported by others (Kim, 1965, 1970; Han and Weitzman, 1970).

Abramson and Lisker (1972) later studied the phonetic significance of the VOT values from a perceptual viewpoint by giving a continuum of synthetic VOT variants (Lisker and Abramson, 1970) to native speakers for identification as Korean syllables. Their results indicated that there must be another dimension that works with VOT in distinguishing the categories, although the timing of glottal adjustments relative to supraglottal articulation does contribute to the consonant distinctions.

Kagaya (1971) investigated laryngeal gestures in Korean stop production closely in a native speaker using fiberscopic observations. He found that there

---

[+]Faculty of Medicine, University of Tokyo, Japan; Haskins Laboratories, New Haven, Conn., 1970-1972.

[++]Department of Linguistics, University of California, San Diego.

[+++]Haskins Laboratories, New Haven, Conn.; on leave from the Faculty of Medicine, University of Toyko, Japan.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

are differences between the stop types in both the time course of glottal width and the apparent glottal conditions in the succeeding vowel segment. In particular, the adjustment of the vocal folds was found to be substantially different for Category I or "forced" type when compared to the other two. In Category I the glottis closed rapidly and there was complete contact of the vocal processes before the onset of voicing, while a slight opening still remained in the membranous portion of the glottis.

Lee and Smith (1971) measured both intraoral and subglottal air pressures simultaneously during the production of the three kinds of Korean stops. They found that subglottal pressure was higher for Category III, the highly aspirated stop, than for the other two categories. They also compared the dynamic patterns of subglottal pressure slope for the three categories and found that the Category III stop showed the most rapid increase in subglottal pressure in the time period immediately before the stop release. They concluded that the highly aspirated stop was the most "dynamic" in this respect.

In recent years, a considerable number of electromyographic (EMG) studies of the laryngeal muscles have been reported. Among those, the Haskins' group (Hirose and Gay, 1972a, 1972b; Hirose, Lisker, and Abramson, 1972) investigated EMG patterns of the intrinsic and extrinsic laryngeal muscles for different kinds of languages and reported that there was a reciprocal pattern of activity between the adductor and abductor muscle groups of the larynx for voiced-voiceless and aspirated-unaspirated contrasts.

The primary purpose of the present study was to investigate electromyographically the actions of the intrinsic laryngeal muscles in production of Korean stop consonants. A native speaker of Korean served as the subject. In a separate experiment an attempt was made to take fiberoptic motion pictures of the glottis of the same subject during stop production.

Method

1. EMG experiment. One of the present authors (C.Y.L.), a native Korean speaker from Kyung-sang-book-do, Sang-ju-goon, Hwa-dog-myun, was the subject in this experiment. He read randomized lists of test sentences 16 times each. In each sentence a test word was embedded in the frame /ikəsi ___ ita/ (This is ___). In the first part of the experiment, test words in the form of CV1 with a short, unstressed vowel were used. The consonant (C) was labial, dental, or velar, and the vowel (V) was /i/, /a/, or /u/. About half of the 27 phonemic sequences thus formed for the test words were nonsense syllables, but they did not violate any phonological constraints of the dialect in question. In the second part of the experiment, test words of the form VCV1 were used.[1]

EMG recordings were made using hooked-wire electrodes. The electrodes were inserted into the interarytenoid muscle (INT) perorally by indirect laryngoscopy using a curved probe, but a percutaneous approach was employed for insertion

[1] It was noted in listening tests and oscillographic observations that the vowel /i/ after /s/ in the frame sentence /ikasi/ was consistently devoiced by this subject, yielding [igəsi̥] as the pronunciation.

into the vocalis (VOC)[2], the lateral cricoarytenoid (LCA), and the cricothyroid (CT) muscles. Insertion into the posterior cricoarytenoid (PCA) was also attempted perorally but proved unsuccessful because of anatomical difficulty. EMG activities from the orbicularis oris (OO) and the sternohyoid (SH)[3] were also recorded percutaneously. A more detailed description of the electrode preparation and insertion techniques can be found in previous reports (Hirose, 1971a; Hirose, Gay, and Strom, 1971).

EMG signals were recorded on a multichannel data recorder simultaneously with acoustic signals and automatic timing markers. The signals were then reproduced and fed into a computer after appropriate rectification and integration. The EMG signal from each electrode pair was averaged over more than 14 selected utterances of each test sentence with reference to a line-up point on the time axis representing a predetermined speech event. In the present experiment, the release of stop closure in each test word was used for the line-up. The data-recording and computer-processing systems used in the present experiment are described in more detail by Port (1971).

2. Fiberoptic observation. Separately from the EMG experiment, motion pictures of the glottis of the same subject were taken using a fiberscope at a film rate of 60 frames per second. As the test utterances isolated nonsense /CVCV/ and /VCV/ words were used, where /V/ was always /i/. Appropriate frame sequences for each type of stop were then examined frame-by-frame with special reference to the time course of glottal width as measured at the vocal processes.

Results

Figure 1 illustrates averaged EMG curves of VOC and CT for the three different bilabial stops in word-initial position. The zero on the abscissa marks the line-up point for averaging, which corresponds to the release of the stop closure. The time axis is marked off every 100 msec. In each type three curves are superimposed for the postconsonantal vowels /i/, /a/, and /u/, which are represented respectively by thin, thick, and dashed lines.

We note in Figure 1 that the curves are similar for a given stop type, i.e., those for different postconsonantal vowels coincide fairly well. This holds true both for VOC and CT as shown in the figure, and also for INT and LCA, which are not shown here.

Figure 1 also shows that the pattern of CT activity is more or less constant for the three stop types, being characterized by two peaks separated by a temporal suppression in the middle portion of the test utterance.

Figure 2 compares the activity of INT and VOC for the test utterances containing the three types of bilabial stops in word-initial position followed by the vowel /i/. The INT activity starts to increase before initiation of the test utterance and, after reaching its peak near the beginning of the first vowel [i],

---

[2]Recordings from VOC were not obtained in the session using VCV1 type test words.

[3]The data for SH will not be discussed in this report.

Figure 1

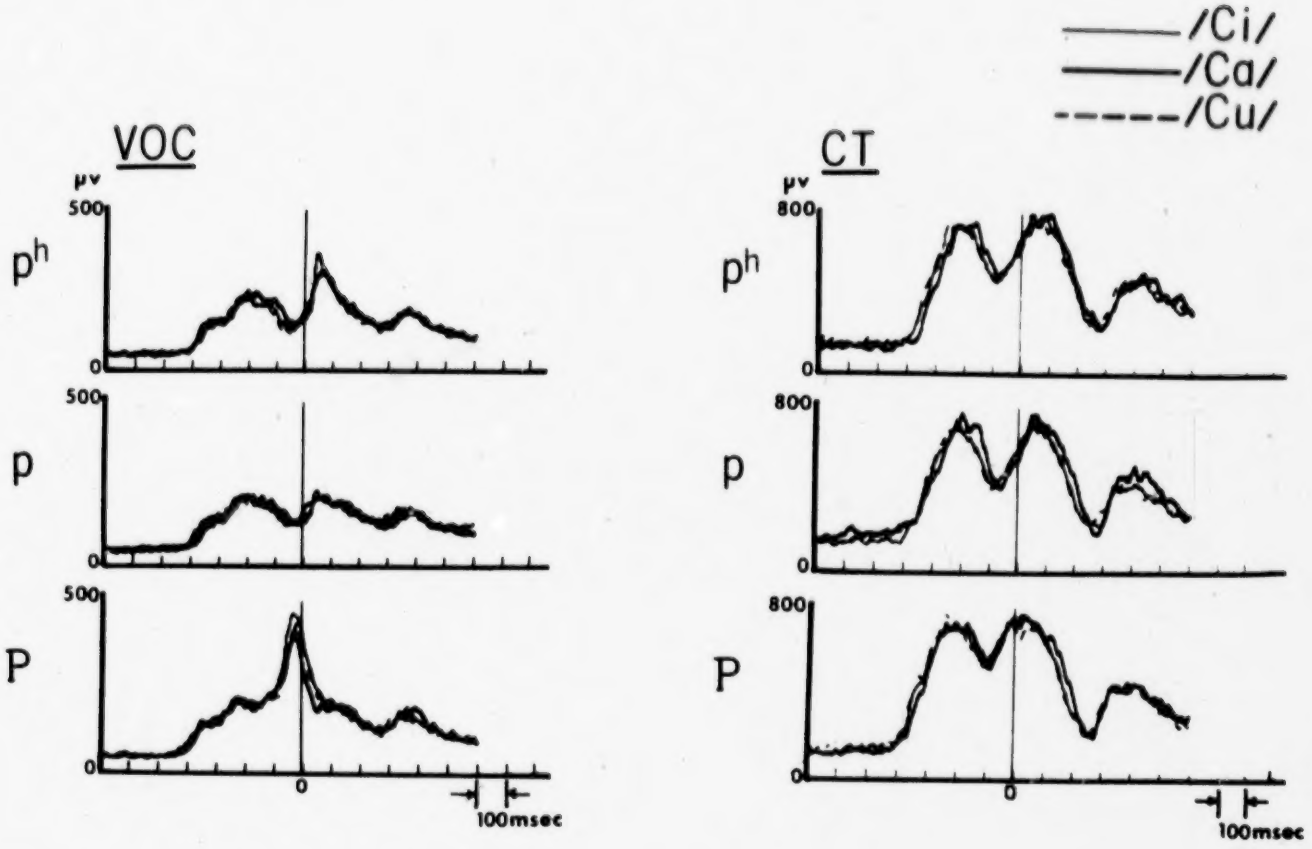


Figure 1: Averaged EMG curves of VOC (left) and CT (right) for the three bilabial stops in word-initial position. In each type, three curves are superimposed for the postconsonantal vowels /i/, /a/, and /u/, represented respectively by thin, thick, and dashed lines. The zero on the abscissa marks the line-up, which corresponds to the release of the stop closure.

Figure 2

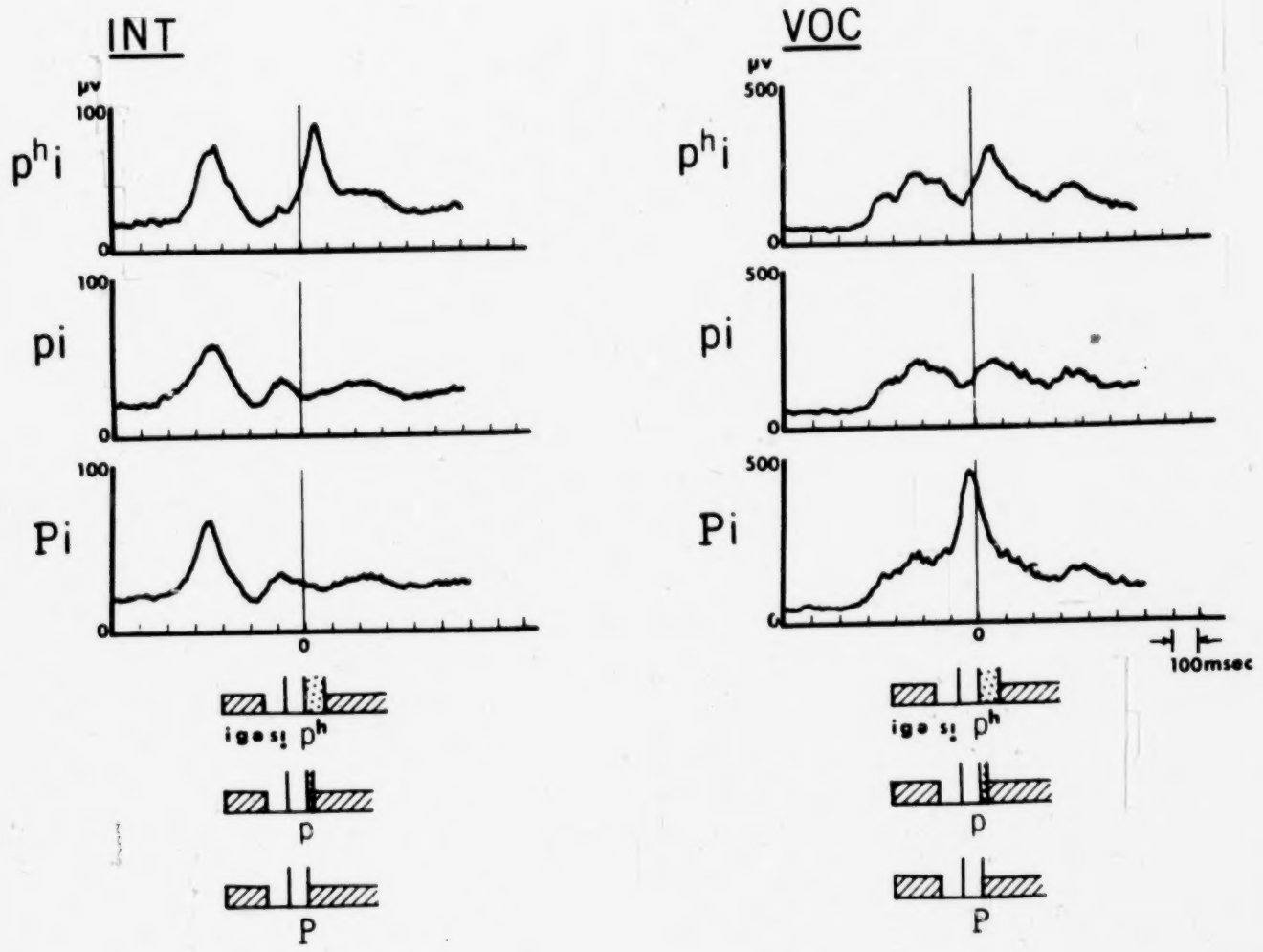


Figure 2: Averaged EMG curves of INT (left) and VOC (right) for the three bilabial stops in word-initial position. The postconsonantal vowel is /i/ for all cases. Timings of speech events are given below the graphs; striped areas represent voiced segments, open areas represent voiceless segments, and dotted areas represent the period of aspiration.

the activity decreases for the voiceless segment [s̆i̥]. In the case of Category III /p^h/, INT activity continues to be suppressed and then steeply increases again near the stop release. In Category II /p/ and Category I /P/, INT activity shows a slight increase after a marked suppression for the [s̆i̥] segment and then stays on a moderate level.

The activity of VOC also starts to increase before the initiation of the utterance but there is a slight delay in timing when compared with that of INT. The pattern for /p^h/ and /p/ is characterized by two peaks, with suppression between the peaks possibly reflecting the voice cessation around the word boundary. For /P/, on the other hand, VOC shows a marked increase in activity immediately before the release.

Figure 3 illustrates the patterns of INT, LCA, and OO activities for the three bilabial stops in word-initial position (left) and in word-medial position (right). The postconsonantal vowel is /i/ for all cases. For the test words with the stop consonant in word-medial position, the onset of the phonated vowel segment at the beginning of the word after the devoiced vowel of the carrier was taken as the line-up point.

The general pattern of LCA activity is similar to that of VOC in that LCA also shows increasing activity before the stop release in the case of /P/, regardless of the position of the consonant, while it shows two separate peaks for both /p^h/ and /p/. There is no discernible difference in the pattern of OO activity among the three different stop types when the consonants are in word-initial position. In word-medial position, however, OO activity is definitely less for Category II /p/, here pronounced [b], than for the other two.

The activity of INT for test utterances with the stop consonants in word-medial position increases before the initiation of the utterance and shows a peak approximately 300 msec before the line-up point, followed by a steep decline appropriate for voiceless [s̆]. The activity increases again approximately 100 msec before the line-up point probably for the vowel segment that precedes the stop closure period and, after reaching the second peak near the line-up, it is then suppressed for the consonantal segment. The suppression is most marked for /p^h/ in both degree and duration. For /p^h/, there is a steep elevation of activity after the period of suppression. For /P/, INT suppression reaches its greatest point earlier than for /p^h/, and is followed by a slight elevation toward a moderate level of activity. For /p/, which is voiced in word-medial position, INT activity gradually decreases after the peak near the line-up point and then sustains a moderate level of activity.

In the case of LCA, the pattern for /p^h/ also appears to be characterized by a marked suppression followed by a steep increase. For /p/ in word-medial position, LCA activity stays moderate for the consonantal segment as well as for the subsequent portion of the test utterance. There is a definite increase in LCA activity for /P/ in word-medial position approximately 150 msec after the line-up, corresponding roughly to the stop closure period.

Figure 4 shows time courses of the glottal width for representative utterance samples by the same subject of the three bilabial stops in absolute initial position. The rectangles represent /P/ (Category I), the circles /p/ (Category II), and the triangles /p^h/ (Category III). Filled rectangles and circles

Figure 3

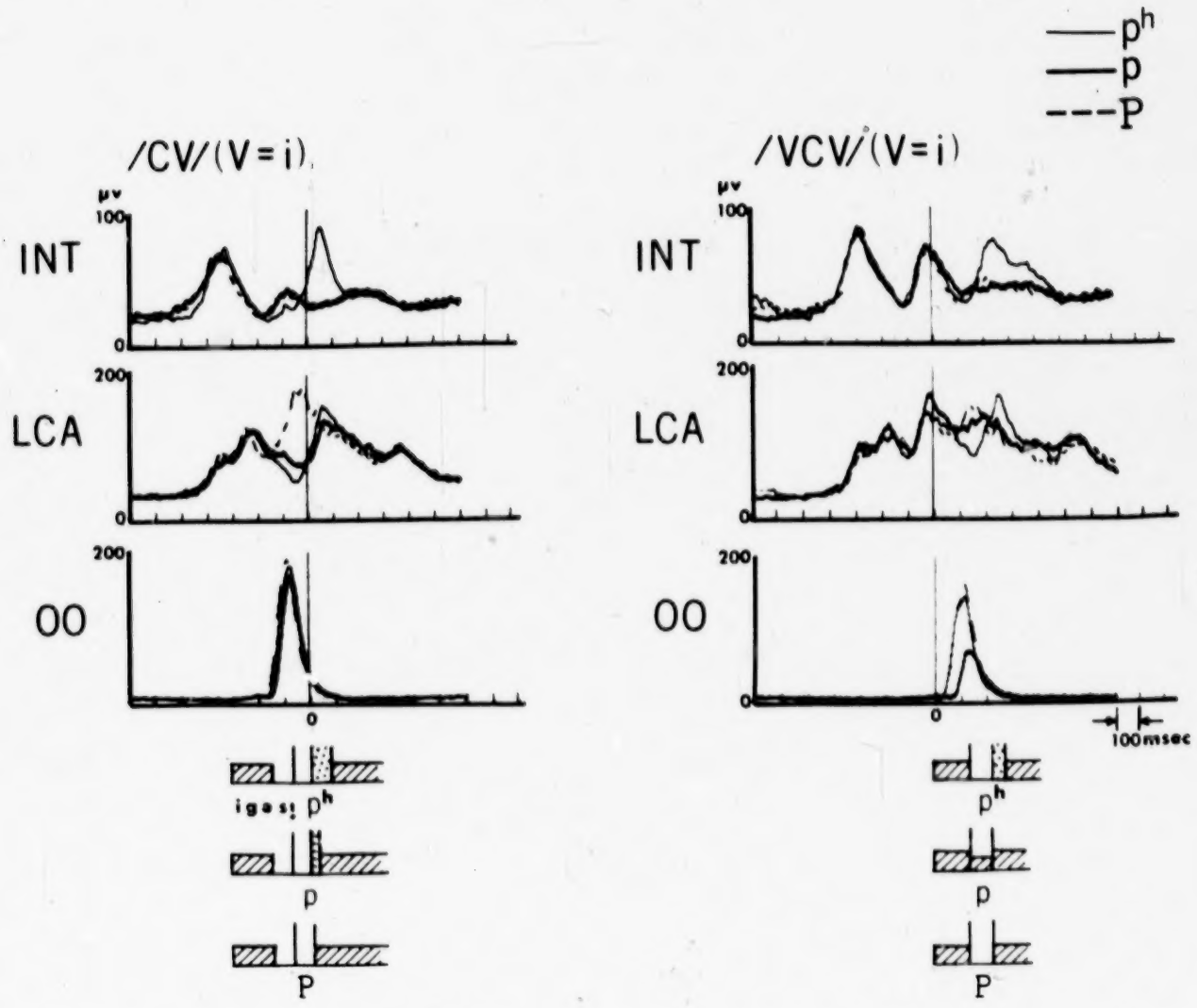


Figure 3: Averaged EMG curves of INT, LCA, and OO for the three bilabial stops in word-initial position (left) and in word-medial position (right). For test words with the stop consonant in word-medial position, the onset of the vowel segment after [s] in the carrier is taken as the lineup.

Figure 4

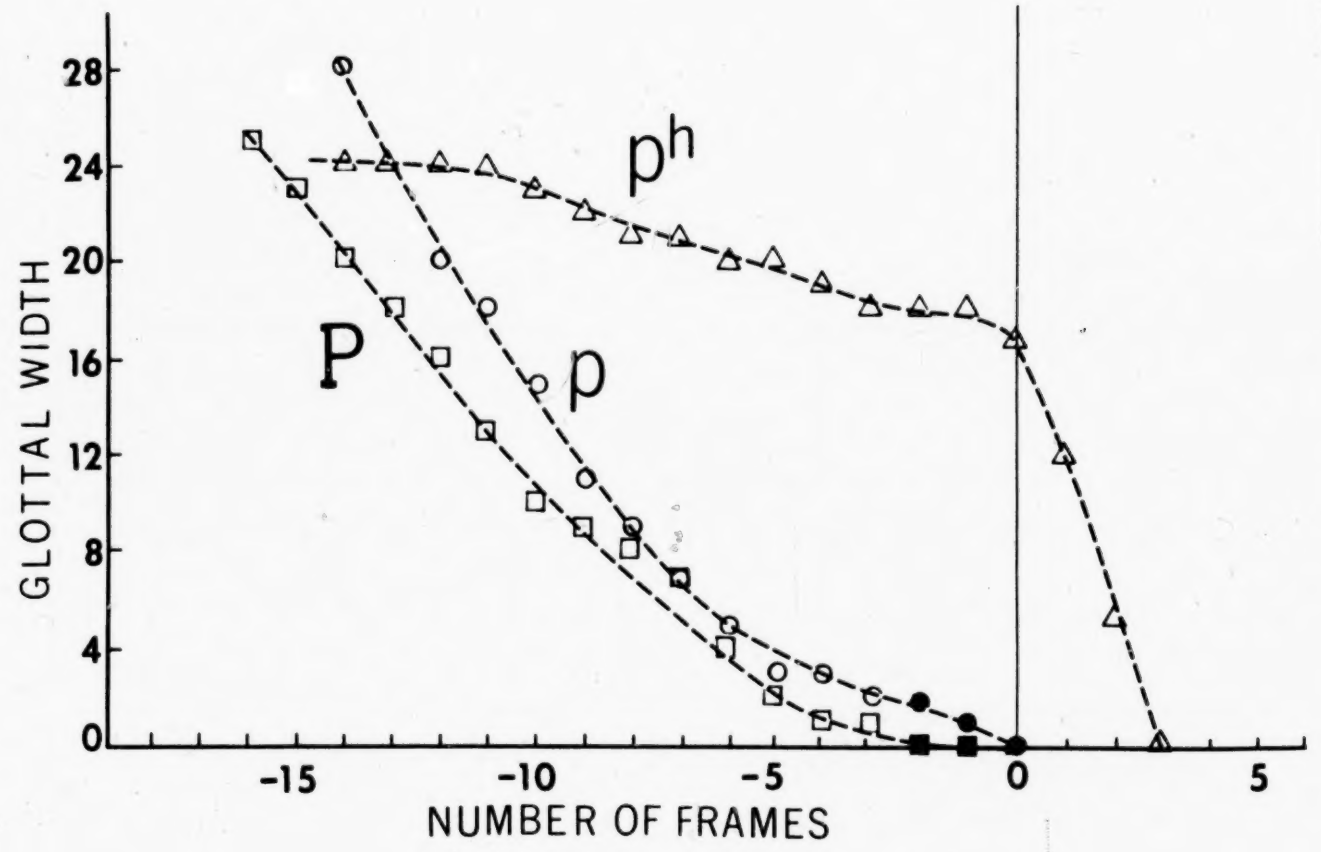


Figure 4: Time courses of the glottal width for representative utterance samples of the three bilabial stops in absolute word-initial position. The rectangles represent /P/ (Category I), the circles /p/ (Category II), and the triangles /p^h/ (Category III). Filled rectangles and circles indicate that vocal fold vibration was observed in that frame. The zero on the abscissa marks the release of stop closure. The film was taken at a rate of 60 fps.

indicate that vocal fold vibration was observed in that frame. The zero on the abscissa marks the release of stop closure.

The figure shows that the glottis begins to close earlier relative to the stop release in Categories I and II, while it stays wide open until the release in Category III. In other words, it seems that there is a considerable difference in glottal width during the consonantal closure period between Category III and the other two. When we compare Category I and Category II, it appears that in Category I the glottis closes somewhat more rapidly and a complete contact of the vocal processes is found before the stop release, while in Category II the glottis closes gradually.

The results for dental and velar stops were essentially comparable to those obtained for bilabial stops in both EMG and fiberoptic experiments.

Discussion

The experimental results of the present study clearly suggest that coordinated actions of the laryngeal muscles characterize the different types of Korean stops.

The "aspirated" stop (Category III) appeared to be characterized by suppression of all the adductor muscles of the larynx immediately preceding the articulatory release. This suppression was always followed by a steep increase in activity which seemed to correspond to the rapid closure of the glottis after stop release, as noted in fiberoptic observations both in this study and elsewhere (Kagaya, 1971).

The pattern of INT activity was almost the same for Category I and Category II stops in word-initial position. It has been observed in previous studies that INT actively participates in the adduction of the vocal fold in speech articulation (Hirose, 1971b; Hirose and Gay, 1972a). The pattern of INT activity is usually known to be reciprocal with that of PCA, the only known abductor of the vocal fold not examined in the present study. In the phonetic environment examined here, INT activity was found to be markedly suppressed for the voiceless segments of [s̆i] (where the glottis seemed to be wide open) after an initial increase for the voiced segment in the preceding cortext. The glottal width during the stop closure period has been found to be narrower in Category I and II stops than in Category III. In the light of this fiberscopic finding it is expected that in the case of Categories I and II the glottal width during the stop closure becomes narrower than for the preceding voiceless [s̆i]. A slight increase in INT activity observed in Categories I and II immediately before the onset of the articulatory closure seems to indicate the active narrowing of the glottis described above.

The patterns of VOC and LCA activity were most characteristic for Category I. Both muscles, VOC in particular, showed a marked increase in activity before the stop release in Category I, which presumably resulted in an increase in inner tension of the vocal folds as well as in constriction of the glottis during or immediately after the articulatory closure. It should be reasonable to assume that these activity patterns of VOC and LCA are the physiological correlates of the Category I stop associated with the subjective impression--possibly including laryngeal sensations in production--of "laryngealization" or

"glottalization" which has often been claimed for this type of stop (Abramson and Lisker, 1972; Ladefoged, 1973). On the basis of fiberoptic and acoustic data, Fujimura (1972) stated that Category I stops were expected to show a marked activity of VOC. The present EMG result seems to support this prediction. We cannot be certain, however, that these findings on VOC and LCA activity in Category I stops should be taken as physiological evidence of so-called "tenseness" of Category I.

It is quite evident, at any rate, that the patterns of VOC and LCA activity are different from that of INT in the production of Category I stops. Our previous studies indicated that the pattern of VOC and/or LCA activity often differed from that of INT in laryngeal articulatory adjustments. For example, LCA and VOC always show marked activity for glottal stop production, while INT does not (Hirose and Gay, 1972b). In our preliminary EMG experiment on Danish subjects, VOC and LCA usually showed a marked increase in activity for the production of Danish stød, while INT did not show any activity related to the stød production. In the light of these findings, it seems reasonable to assume that VOC and LCA play a different role from INT in certain types of laryngeal adjustments. In other words, it can be assumed that there is a functional differentiation of the adductor muscles of the larynx, although INT, VOC, and LCA are often grouped together as adductor muscles in the classical sense.

It is also interesting to note that the pattern of OO activity was different between word-initial and medial positions; i.e., it was markedly low for Category II stops in word-medial position, while it was almost the same for all the stop types in word-initial position. One may argue that the lower OO activity for Category II stops in word-medial position could be related to the so-called "laxness." The exact nature of the "tense-lax" feature has not been well documented. In particular, its physiological correlates are still ambiguous, although there have been several reports claiming that tenseness exists in reality in terms of overall tensing of the speech muscles or of a stronger organic pressure (Fischer-Jørgensen, 1968; Malécot, 1970). In any event, it should be stressed that this difference in OO activity is observed only in word-medial position where the occlusion of a Category II stop is completely voiced; in word-initial position OO activity is not distinctive. It is inappropriate at this point to come to a conclusion about the reality of "tense-lax" opposition as a universal feature in many different languages. However, at least for Korean stops, the laryngeal articulatory adjustment is not limited in a simple dimension of adduction-abduction of the vocal folds: another dimension, represented by VOC activity for example, also must be taken into consideration.

REFERENCES

Abramson, A. S. and L. Lisker. (1972) Voice timing in Korean stops. Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal, 1971 (The Hague: Mouton) 439-446.

Fischer-Jørgensen, E. (1968) Voicing, tenseness, and aspiration in stop consonants, with special reference to French and Danish. Annual Report, Institute of Phonetics, University of Copenhagen 3, 63-114.

Fujimura, O. (1972) Acoustics of Speech: Speech and Cortical Functioning. (New York: Academic Press) 107-165.

Han, M. S. and R. S. Weitzman. (1970) Acoustic features of Korean /P,T,K/, /p,t,k/ and /p^h,t^h,k^h/. Phonetica 22, 112-128.

Hirose, H. (1971a) Electromyography of the articulatory muscles: Current instrumentation and technique. Haskins Laboratories Status Report on Speech Research SR-25/26, 73-86.

Hirose, H. (1971b) Laryngeal adjustments for vowel devoicing in Japanese. Haskins Laboratories Status Report on Speech Research SR-28, 157-166.

Hirose, H. and T. Gay. (1972a) The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study. Phonetica 25, 140-164.

Hirose, H. and T. Gay. (1972b) Laryngeal control in vocal attack: An electromyographic study. Haskins Laboratories Status Report on Speech Research SR-29/30, 49-60.

Hirose, H., T. Gay, and M. Strome. (1971) Electrode insertion techniques for laryngeal electromyography. J. Acoust. Soc. Amer. 50, 1449-1450.

Hirose, H., L. Lisker, and A. S. Abramson. (1972) Physiological aspects of certain laryngeal features in stop production. Haskins Laboratories Status Report on Speech Research SR-31/32, 183-191.

Kagaya, R. (1971) Laryngeal gestures in Korean stop consonants. Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo 5, 15-24.

Kim, C. (1965) On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). Word 21, 339-359.

Kim, C. (1970) A theory of aspiration. Phonetica 21, 107-116.

Ladefoged, P. (1973) The features of the larynx. J. Phonetics 1, 73-83.

Lee, C. Y. and S. S. Smith. (1971) A study of subglottal air pressure in Korean stop consonants. (Preliminary version, presented at the 82nd meeting of the Acoustical Society of America, Denver, Colo.)

Lisker, L. and A. S. Abramson. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. Word 20, 384-422.

Lisker, L. and A. S. Abramson. (1970) The voicing dimension: Some experiments in comparative phonetics. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. (Prague: Academia) 563-567.

Malécot, A. (1970) The lenis-fortis opposition: Its physiological parameters. J. Acoust. Soc. Amer. 47, 1588-1592.

Martin, S. E. (1951) Korean phonemics. Language 27, 519-533.

Port, D. K. (1971) The EMG data system. Haskins Laboratories Status Report on Speech Research SR-25/26, 67-72.

Umeda, H. and N. Umeda. (1965) Acoustical features of Korean "forced" consonants. Gengo Kenkyu [J. Ling. Soc. Japan] 48, 23-33.

Patterns of Palatoglossus Activity and Their Implications for Speech Organization*

F. Bell-Berti[+] and H. Hirose[++]

While the levator palatini has been established as the muscle primarily responsible for soft palate elevation, no antagonist muscle has been found to be responsible for soft palate lowering. Some investigators (Fritzell, 1969; Lubker, Fritzell, and Lindquist, 1970) have suggested that the palatoglossus is a muscle serving this palate-lowering function. Other investigators (Berti and Hirose, 1971), who have not found evidence supporting this hypothesis, have found instead that palatoglossus activity corresponds to tongue body movements: velar consonant and back vowel articulations.

We shall report on electromyographic (EMG) data obtained from two subjects: the first, LJR, is a native speaker of American English; the second subject, BG, is a native speaker of Swedish and is the same subject whose data were reported by Lubker, Fritzell, and Lindquist (1970) in a study of nasal articulation in Swedish. The test utterances were nonsense disyllables designed to determine the effect of vowel color, and the place and manner of stop consonant articulation on palatoglossus activity. For example, one utterance is /fapmap/. The data were processed using the Haskins Laboratories' EMG system.

Results and Discussion

We will begin our discussion by examining the EMG potentials associated with labial nasal articulations.

The zero point in all the figures occurs at the acoustic boundary between the oral and nasal stops. The acoustic signals are represented above each figure. Of course, the EMG signal associated with an acoustic event precedes that event.

---

*Paper presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., April 1973.

[+]Haskins Laboratories, New Haven, Conn.; Montclair State College, Upper Montclair, N. J.; and The Graduate School, The City University of New York.

[++]Faculty of Medicine, University of Tokyo, Japan. Visiting researcher, Haskins Laboratories, 1970-1972.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

There is no palatoglossus activity associated with labial nasal articulation for subject LJR--no activity is seen for /fimpip/ or /fipmip/ (Figure 1). Subject BG, whose data are displayed in the lower half of Figure 1, does show palatoglossus activity for labial nasal articulation. For subject BG, the activity peak occurs earlier in /fimpip/ (-200 msec) than in /fipmip/ (-110 msec): the peak shifts in the direction of the nasal, occurring earlier for /fimpip/ than for /fipmip/.

No clear activity is observed for any of the vowels for subject LJR (Figure 2). Subject BG again presents activity for the labial nasal in /fimpip/ (Figure 2). In addition, activity is evident for the vowels in /fɑmpɑp/ and /fumpup/. The greatest activity occurs for both of the vowels in /fumpup/.

Peaks are observed for the velar oral stop in /fɑkmɑp/ near -250 msec for both subjects (Figure 3). In addition, subject BG has a second peak at -70 msec for the labial nasal, /m/, in /fɑkmɑp/. Both subjects show peaks at -180 msec for the velar nasal, /ŋ/, in fɑŋpɑp/.

In summary, palatoglossus activity for subject LJR corresponds essentially to velar articulations. Activity is observed only when the stop is velar--regardless of the oral or nasal manner of articulation. Subject BG, however, presents palatoglossus activity for all nasal articulations. He also shows activity for back vowels, with greater activity for /u/ than /ɑ/. Subject BG also shows activity associated with velar articulations: /k/ and /ŋ/.

We have concluded, therefore, that palatoglossus activity is primarily associated with tongue body movements, but may be implicated in the nasal manner of articulation in some speakers. We may not yet specify whether these differences in palatoglossus function (i.e., tongue body vs. nasal gestures)[1] are language-specific or idiosyncratic in nature: our Swedish speaker was the same individual whose data were reported by Lubker, Fritzell, and Lindquist (1970). We await further cross-language studies to determine the cause of these differences. We may say, though, that no universal mode of nasal articulation, corresponding to the universal mode of oral articulation found in levator palatini function, may be specified: that is, while palatoglossus function for nasal articulation may exist for speakers of some languages, this function does not occur for all speakers of all languages.

Subject BG, the Swedish speaker of this experiment, shows palatoglossus activity for vowels, nasals, and velar consonants. One additional finding for this subject, BG, is of interest.

When a consonant with palatoglossus activity follows a vowel with palatoglossus activity, the two peaks merge and form one broader peak (the /-um-/ in /fumpup/) (Figure 4). When this consonant precedes a vowel with palatoglossus activity, two separate peaks are observed, the /-mu-/ of /fupmup/ (the dotted line). The peaks in /fumpup/ do not merge as a consequence of a time-smear effect: the nasal in /-mu-/ (/fupmup/) is shorter than the first vowel in

---

[1] Data from other speakers of American English (Bell-Berti, 1973) show palatoglossus activity only for vowels.

palatoglossus

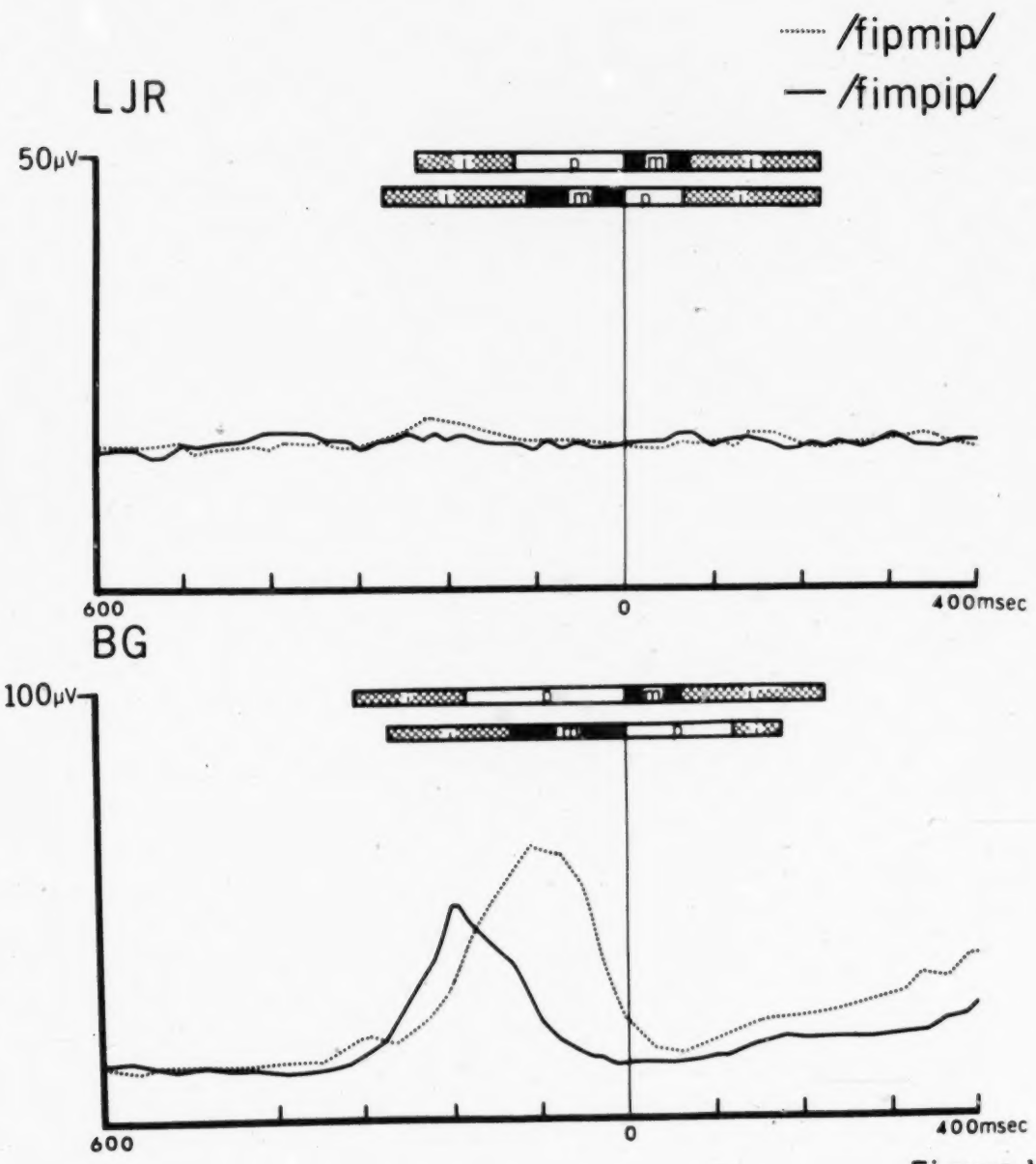


Figure 1

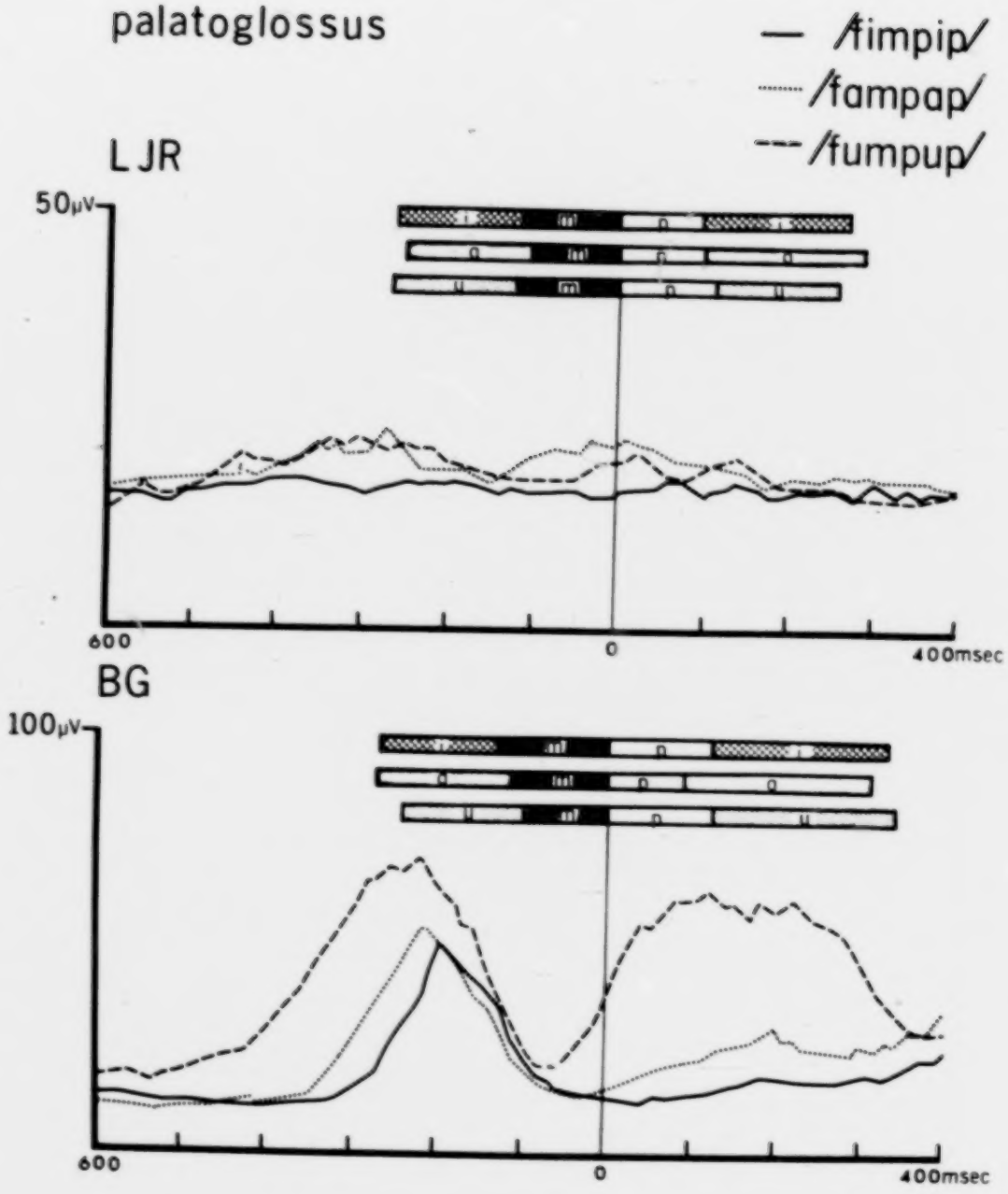

palatoglossus
— /ɪimpip/
····· /ɪampap/
--- /ɪumpup/
LJR
50μV
600
0
400msec
BG
100μV
600
0
400msec

Figure 2

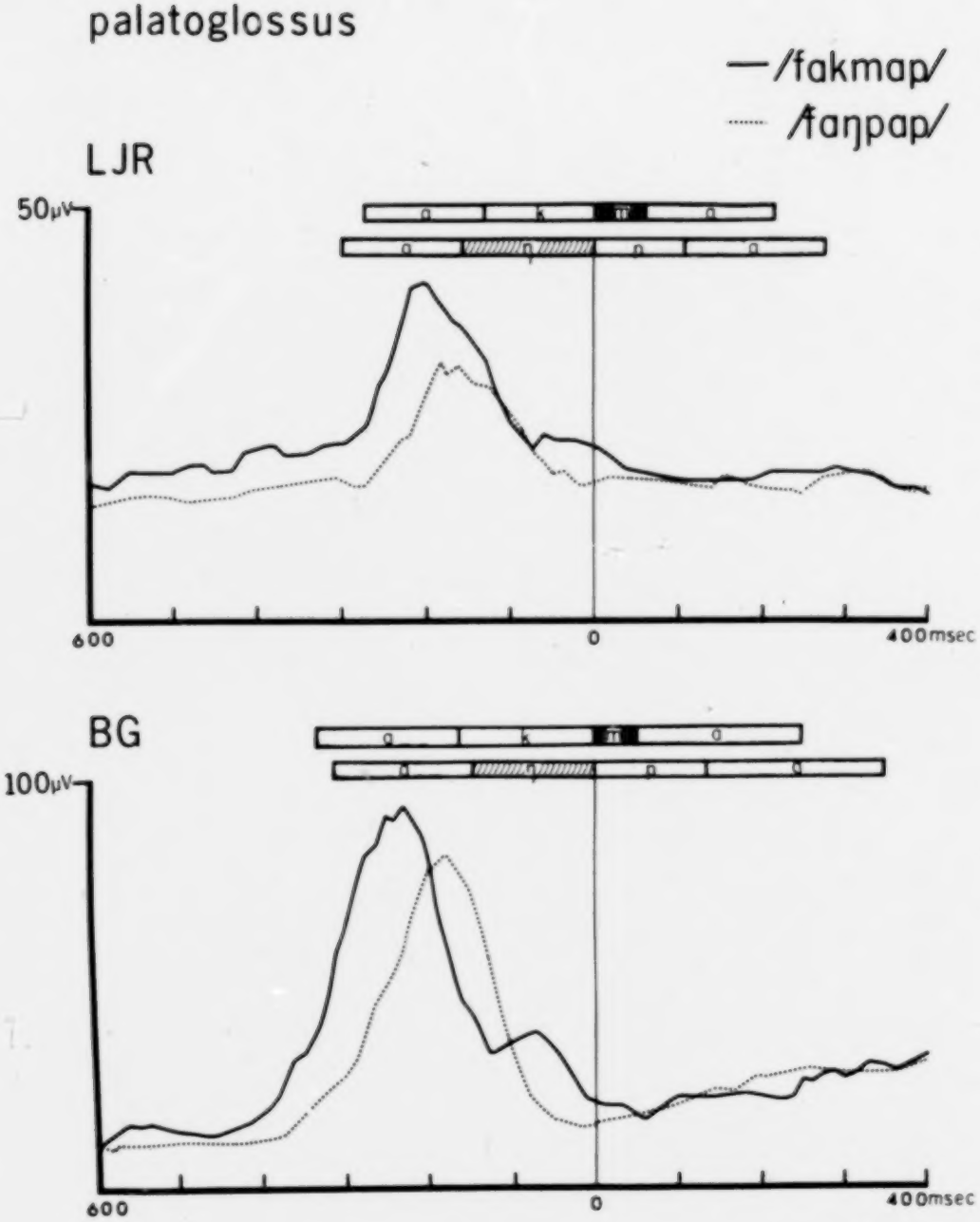

palatoglossus
—/fakmap/
---- /faŋpap/
LJR
50μV
600
0
400msec
BG
100μV
600
0
400msec

Figure 3

palatoglossus

— /ʌumpup/

····· /ʌupmup/

BG

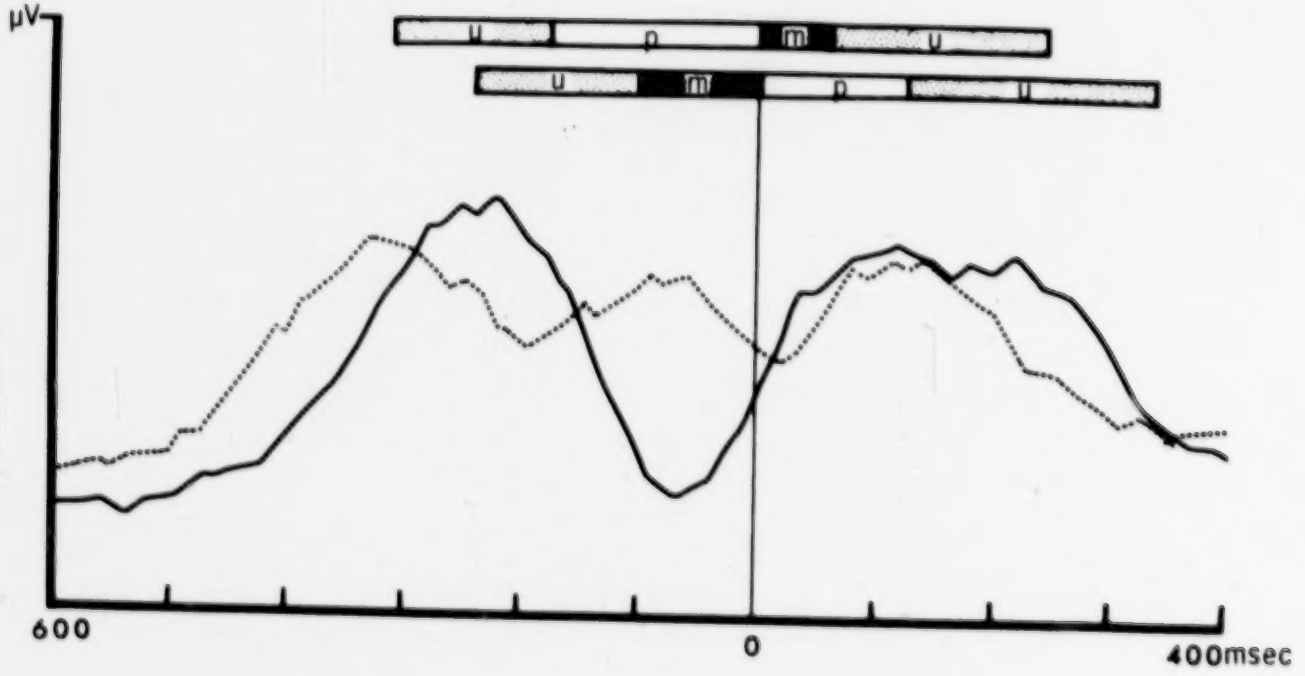


Figure 4

/-um-/ (fumpup). If the effect were due to temporal overlap of the EMG signals, the peaks would merge in /-mu-/ utterances, where the beginnings of the involved phones are closer than they are in /-um-/ utterances.

Two possible explanations emerge for this difference in the pattern of palatoglossus activity in vowel-nasal and nasal-vowel combinations. The first is that momentary relaxation of the palatoglossus is required to facilitate initiation of palatal elevation by the levator palatini for the production of an oral /u/, an articulation requiring a fairly tight velopharyngeal seal.

Another more tantalizing, but highly speculative, explanation is that this pattern is a reflection of some aspect of syllabic organization: that motor commands may be merged for VC sequences but not for CV sequences. This speculation is based on limited but highly reliable data: the pattern occurs for all cases having palatoglossus activity for both the vowel and consonant members of the CV and VC pairs (including oral velar stops). It may reflect the generally greater constriction of the oral cavity for consonants than for vowels: activity may increase through a vowel into a consonant but must decrease to permit a reduction of oral cavity constriction for a vowel following a consonant.

While no final statement may be made about the cause of this difference in activity patterns, the observation warrants further investigation.

REFERENCES

Bell-Berti, F. (1973) The velopharyngeal mechanism: An electromyographic study. Unpublished Ph.D. thesis, The City University of New York.

Berti, F. and H. Hirose. (1971) Velopharyngeal function in oral/nasal articulation and voicing gestures. Haskins Laboratories Status Report on Speech Research SR-28, 143-156.

Fritzell, B. (1969) The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. Acta Otolaryng. Suppl. 250.

Lubker, J. F., B. Fritzell, and J. Lindquist. (1970) Velopharyngeal function: An electromyographic study. Speech Transmission Laboratory Quarterly Progress and Status Report (Stockholm, Sweden: Royal Institute of Technology) STL-QPSR 4/1970, 9-20.

ABSTRACT

Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency*

James E. Atkinson+

In a study of intonation in American English several acoustic and physiological correlates of the fundamental voice frequency (Fo) were investigated. The goal was to determine the linguistically relevant acoustic and physiological aspects of Fo and to relate these within a unified, phonetic feature theory. Several apsects of Fo production were studied.

Inter- and Intra-Speaker Fo Variability

A measure of the amount of Fo variability was obtained for various repetitions by a single speaker and compared with that from several different speakers. The results show the inter- and intra-speaker variabilities to be of the same order of magnitude. A detailed study of the type of variability and its probable causes was presented to determine its perceptual relevance. The results offer strong evidence that nothing finer than a binary distinction (± Prominence) can be made in terms of Fo. Phonetic theories which demand many fine Fo distinctions seem to be overspecified.

Physiological Factors

An electromyographic (EMG) study of several laryngeal muscles was conducted using hooked-wire electrodes (Hirose, 1971) and the Haskins Laboratories' EMG data system described by Port (1971). The muscles investigated were: vocalis, cricothyroid, lateral cricoarytenoid, sternothyroid, and sternohyoid. These were studied for several types of sentence intonation.

Subglottal, transglottal, and oral air pressure were measured for the same utterances. Subglottal pressure was obtained using a tracheal catheter. In

---

*Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Connecticut, Storrs.

+Naval Underwater Systems Center, New London, Conn.

Acknowledgment: Partial support for this research came from the Naval Underwater Systems Center, New London, Conn., and from grants to Haskins Laboratories, New Haven, Conn.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

addition, lung volume and air-flow rates were obtained for these utterances. In this way a detailed study of all of the physiological factors known to affect Fo was possible.

Intrinsic Fo in Vowels

The above physiological measures were also obtained for various steady-state vowels in an attempt to explain the phenomenon of intrinsic Fo differences between vowels. Traditionally this has been explained in terms of mechanical interaction between the tongue and larynx, although recent cineradiographic evidence contradicts this. The results of this study supported an explanation offered by Flanagan and Landgraf (1968) which accounted for these differences in terms of source-system acoustic coupling effects.

Computer-Implemented Correlation Analysis

A correlation analysis technique was developed to allow a detailed look at the physiological factors controlling Fo and at their interactions. The results indicated that various factors may be involved in controlling Fo depending upon the type of utterances, and the results suggested a "modal theory" of laryngeal control employing two different laryngeal states, which appeared to be mediated by the sternohyoid muscle in this study.

Taken as a whole, the results of this study show an essential interaction between the physical constraints and capabilities of the vocal apparatus and the prosodic features. These features (Prominence and Breath-group) seem to be structured and implemented to take maximum advantage of the normal vegetative process of respiration, and to minimize the number of additional adjustments from this state. Most simple declarative statements follow this pattern (denoted "-Breath-group"), and the single most important factor in controlling Fo for these utterances is subglottal pressure, although laryngeal adjustments also may play a role. The evidence shows that utterances like yes-no questions (denoted "+Breath-group") are "marked" in the sense that special respiratory and laryngeal adjustments are required. All of this supports the notion that the fundamental unit of intonation is the breath-group. Its function is to help segment the nearly continuous train of speech sounds into linguistic units, and to denote certain features of the underlying constituent structure. The phonetic/phonological features (Prominence and Breath-group) are "signaling units" which can be implemented in various ways. The results show that just as the segmental phonemes are encoded into syllable-sized units (see Mattingly and Liberman, 1969; Liberman, 1970), so the prosodic features are encoded and must be perceived in terms of the entire breath-group, and not as discrete sequential features. This study supports the notions of a motor theory of speech perception (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967) and of an archetypal breath-group as suggested by Lieberman (1967).

REFERENCES

Flanagan, J. L. and L. Landgraf. (1968) Self-oscillating source for vocal tract synthesizers. IEEE Trans. Audio 16, 57-64.

Hirose, H. (1971) Electromyography of the articulatory muscles: Current instrumentation and technique. Haskins Laboratories Status Report on Speech Research SR-25/26, 73-86.

Liberman, A. (1970) The grammars of speech and language. Cog. Psychol. 1, 301-323.

Liberman, A., F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. (1967) Perception of the speech code. Psychol. Rev. 74, 431-461.

Lieberman, P. (1967) Intonation, Perception and Language. (Cambridge, Mass.: MIT Press).

Mattingly, I. G. and A. Liberman. (1969) The speech code and the physiology of language. In Information Processing in the Nervous System, ed. by K. N. Leibovic. (New York: Springer-Verlag).

Port, D. K. (1971) The EMG data system. Haskins Laboratories Status Report on Speech Research SR-25/26, 67-72.

ABSTRACT

Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses*

Charles C. Wood+

The relation between an acoustic speech signal and its phonetic message appears to be a complex and highly efficient code, based on parallel transmission of phonetic information in the speech signal. Previous experiments have suggested that the perception of this "speech code" involves specialized phonetic decoding mechanisms located in the dominant cerebral hemisphere, mechanisms that are not involved in the perception of nonspeech sounds. This suggestion has received additional support from the demonstration that some components of a speech signal require specialized phonetic processing for their perception, while other components can be processed by the general auditory system alone. For example, recent experiments by the author have shown that different levels of processing underlie the perception of auditory and phonetic dimensions of synthetic speech stimuli. In one experiment, reaction time (RT) data indicated that while the auditory dimension could be processed independently of the phonetic dimension, the phonetic dimension could not be processed independently of the auditory dimension. In a second experiment, averaged evoked potentials were recorded during the processing of the same auditory and phonetic dimensions.

---

*A dissertation presented to the faculty of the Graduate School of Yale University, New Haven, Conn., in candidacy for the degree of Doctor of Philosophy, June 1973. This dissertation was done in the Department of Psychology under the direction of W. R. Goff, R. S. Day, and W. R. Garner. The experimental work was carried out for the most part in the Neuropsychology Laboratory of the West Haven Veterans Administration Hospital; the synthetic speech stimuli were generated and analyzed in the Haskins Laboratories. The thesis is summarized here primarily because of its relevance to the speech research program of Haskins Laboratories. The content of the thesis will appear in the next regular issue of these Status Reports.

+Now at the Walter Reed Army Institute of Research, Department of Experimental Psychophysiology, Walter Reed Army Medical Center, Washington, D. C.

Acknowledgment: This dissertation was supported in part by the Veterans Administration, Public Health Service National Institute of Mental Health Grant M-05286 and National Science Foundation Grants GB-3919 and GB-5782 to W. R. Goff, and by a National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-34 (1973)]

This experiment demonstrated that the processing of the phonetic dimension was accompanied by neural events in the left hemisphere which did not occur during the processing of the auditory dimension. Thus, using very different response measures, both experiments suggest that perception of the phonetic dimension involved an additional level of processing which was not required for the auditory dimension.

The present investigation was designed to substantiate the distinction between auditory and phonetic levels of processing made by the initial experiments, and to provide additional information concerning the nature of the specialized phonetic level. Instead of collecting the RT and evoked potential data separately, the present experiments employed a single paradigm to obtain both sets of data. The first experiment completely replicated the RT and evoked potential results for auditory and phonetic dimensions obtained separately in the initial experiments. The second experiment provided control data demonstrating that the results attributed to the phonetic level did not occur for two auditory dimensions. The third experiment showed that the phonetic level is specialized for the extraction of abstract phonetic features, not for the detection of particular acoustic features in the speech signal. The fourth experiment suggested that while the phonetic level is linguistic in nature, it is not required for the processing of all acoustic dimensions that can convey linguistic information. Additional analyses of the neurophysiological data demonstrated that the evoked potential differences between auditory and phonetic dimensions were not associated with differences in: 1) frequency spectra or amplitude distributions of the background electroencephalogram (EEG); 2) pre-stimulus baseline changes related to the contingent negative variation (CNV); or 3) averaged activity synchronized to subjects' motor responses. Taken together, the RT and evoked potential data of the present experiments provide a strong set of converging operations upon the distinction between auditory and phonetic levels of processing in speech perception, and upon the idea that the phonetic level involves specialized language mechanisms which are lateralized in one cerebral hemisphere.

II. PUBLICATIONS AND REPORTS

III. APPENDIX

PUBLICATIONS AND REPORTS

Publications and Manuscripts

The following three papers appeared in Proceedings of the Seventh International Congress of Phonetic Sciences, Montreal, 1971. (The Hague: Mouton, 1972).

Glottal Modes in Consonant Distinctions.
Leigh Lisker and Arthur S. Abramson, 366-370.

Voice Timing in Korean Stops.
Arthur S. Abramson and Leigh Lisker, 439-446.

Further Experimental Studies of Fundamental Frequency Contours.
Michael Studdert-Kennedy and Kerstin Hadding, 1024-1031.

Phonetic Ability and Related Anatomy of the Newborn and Adult Human, Neanderthal Man, and Chimpanzee. Philip Lieberman, Edmund S. Crelin, and Dennis H. Klatt. American Anthropologist (1972) 74, 287-307.

Silent Articulation. Katherine S. Harris. Science (1972) 176, 1114-1115.

Word-Final Stops in Thai. Arthur S. Abramson. In Tai Phonetics and Phonology, ed. by J. G. Harris and R. B. Noss. (Bangkok, Thailand: Central Institute of English Language, 1973) 1-7.

A Plan for the Field Evaluation of an Automated Reading System for the Blind. P. W. Nye, J. D. Hankins, T. Rand, I. G. Mattingly, and F. S. Cooper. IEEE Transactions on Audio and Electroacoustics (June 1973) AU-21, 265-268.

Olson's Projective Verse and the Use of Breath Control as a Structural Element. Marcia R. Lieberman and Philip Lieberman. Language and Style (1973) 5, 287-298.

The Phi Coefficient as an Index of Ear Differences in Dichotic Listening. Gary M. Kuhn. Cortex (in press).

Hemiretinae and Nonmonotonic Masking Functions with Overlapping Stimuli. Claire Farley Michaels and M. T. Turvey. Bulletin of the Psychonomic Society (in press).

*Phonological Fusion of Synthetic Stimuli in Dichotic and Binaural Presentation. James E. Cutting.

*Laryngeal Control in Korean Stop Production. Hajime Hirose, Charles Y. Lee, and Tatsujiro Ushijima

*Phonological Fusion of Stimuli Produced by Different Vocal Tracts. James E. Cutting.

---

*Appears in this report, SR-34.

*Hemispheric Specialization for Speech Perception in Six-Year-Old Black and White Children from Low and Middle Socioeconomic Classes. M. F. Dorman and Donna S. Geffner.

*Oral Feedback, Part I: Variability of the Effect of Nerve-Block Anesthesia Upon Speech. Gloria Jones Borden, Katherine S. Harris, and William Oliver.

*Oral Feedback, Part II: An Electromyographic Study of Speech Under Nerve-Block Anesthesia. Gloria Jones Borden, Katherine S. Harris, and Lorne Catena.

Reports and Oral Presentations

*Phonetic Prerequisites for First-Language Acquisition. Ignatius G. Mattingly. Presented at the International Symposium on First-Language Acquisition, Florence, Italy, September 1972; and at the Society for Research in Child Development meeting, Philadelphia, Pa., March 1973.

Phenomena in Cognitive Psychology. Ruth S. Day. Connecticut Junior Science and Humanities Symposium, Yale University, New Haven, Conn., 2 April 1973.

The following seven papers were presented at the 85th meeting of the Acoustical Society of America, Boston, Mass., 10-13 April 1973.

* Reaction Times to Comparisons Within and Across Phonetic Categories: Evidence for Auditory and Phonetic Levels of Processing. David B. Pisoni and Jeffrey Tash.

* The Role of Auditory Short-Term Memory in Vowel Perception. David B. Pisoni

* Effects of Attenuation of One of Two Channels on Perception of Opposing Pairs of Nonsense Syllables when Monotically and Dichotically Presented. (Retitled for SR-34) Susan Brady-Wood and Donald Shankweiler.

* Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects. Ruth S. Day.

* Patterns of Palatoglossus Activity and Their Implications for Speech Organization. F. Bell-Berti and H. Hirose.

Degree of Phrasal Stress: A Stable Lexical Feature? Jane H. Gaitenby, George N. Sholes, and Gary M. Kuhn.

A Two-Pass Procedure for Synthesis-By-Rule. Gary Kuhn.

Electromyographic Studies of Articulatory Organization. K. S. Harris. Invited paper, Symposium on Organized Function in Craniofacial Dynamics, International Organization for Dental Research, Washington, D. C., 12 April 1973.

Evaluation of New Techniques for Research in Oral Muscle Function. K. S. Harris. Invited paper, Research Week, Harvard Dental School, Cambridge, Mass., 24 April 1973.

*On Learning "Secret Languages." Ruth S. Day. Presented at the Eastern Psychological Association meeting, Washington, D. C., 3 May 1973.

Hemispheric Specialization and Individual Differences in Cognition. Ruth S. Day. Invited address, Center for Advanced Study in the Behavioral Sciences, Stanford, Calif., 11 May 1973.

Colloquia. Ruth S. Day. Wesleyan University, Middletown, Conn., 25 April 1973; Stanford University, Stanford, Calif., 11 May 1973.

*A Note on the Relation between Action and Perception. M. T. Turvey. Invited address, Allerton Conference of the North American Society for the Psychology of Sport and Physical Activity, Monticello, Ill., 13-17 May 1973.

Some Aspects of Speech Perception. M. T. Turvey. Hampshire College, Amherst, Mass., May 1973.

Two Central Processes in Vision. M. T. Turvey. Harvard University, Boston, Mass., May 1973.

Invited Talks. M. T. Turvey, R. S. Day, D. B. Pisoni, D. P. Shankweiler, M. Studdert-Kennedy, and J. E. Cutting. Haskins Laboratories' Workshop on Speech Perception, Wallingford, Conn., 5-6 June 1973.

Can the Pigeon See Sideways--A Study of the Visual System. Pat Nye. Yale University, New Haven, Conn., Summer Seminar on Image Processing, 13 June 1973.

Dissertations

*Levels of Processing in Phonological Fusion. James Eric Cutting. Ph.D. dissertation, Yale University, New Haven, Conn., 1973.

Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses. Charles C. Wood. Ph.D. dissertation, Yale University, New Haven, Conn., 1973. (Abstract in SR-34.)

Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Frequency. James E. Atkinson. Ph.D. dissertation, University of Connecticut, Storrs, 1973. (Abstract in SR-34.)

APPENDIX

DDC (Defense Documentation Center) and ERIC (Educational Resources Information Center) numbers:

SR-21/22 to SR-31/32

Status Report		DDC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	
SR-33	January - March 1973		

AD numbers may be ordered from: U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from: ERIC Document Reproduction Service
Leasco Information Products, Inc.
P. O. Drawer O
Bethesda, Maryland 20014

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Haskins Laboratories, Inc. 270 Crown Street New Haven, Connecticut 06510	Unclassified
	2b. GROUP N/A

3. REPORT TITLE

Status Report on Speech Research, No. 34, April-June 1973

4. DESCRIPTIVE NOTES (Type of report and, inclusive dates)
Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)

Staff of Haskins Laboratories; Franklin S. Cooper, P.I.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1973	232	243

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
ONR Contract N00014-67-A-0129-0001 and-0002 NIDR: Grant DE-01774 NICHD: Grant HD-01994 NIH/DRFR: Grant RR-5596 NSF: Grant GS-28354 VA/PSAS Contract V101(134)P-71 NICHD Contract NIH-71-2420 The Seeing Eye, Inc. Equipment Grant	SR-34 (1973)
	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited.*

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
N/A	See No. 8

13. ABSTRACT

This report (1 April-30 June) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

-Levels of Processing in Phonological Fusion
-Phonological Fusion of Synthetic Stimuli
-Phonological Fusion of Stimuli Produced by Different Vocal Tracts
-Phonetic Prerequisites for First-Language Acquisition
-Relation between Action and Perception
-Reaction Times: Evidence for Auditory and Phonetic Levels of Processing
-Auditory Short-Term Memory in Vowel Perception
-Effects of Amplitude Variation on Auditory Rivalry Task
-Digit-Span Memory in Language-Bound and Stimulus-Bound Subjects
-Learning "Secret Languages"
-Hemispheric Specialization for Speech Perception in Six-Year-Old Children
-Effect of Nerve-Block Anesthesia Upon Speech
-Electromyographic Study of Speech Under Nerve-Block Anesthesia
-Laryngeal Control in Korean Stop Production
-Patterns of Palatoglossus Activity--EMG
-Intonation in Speech: Study of Fundamental Frequency
-Levels of Processing in Speech Perception: Neurophysiological and Information-Processing Analyses

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

UNCLASSIFIED
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech perception						
Speech: levels of processing						
Speech: sensory feedback						
Speech: EMG and sensory feedback						
Speech: nasal articulation						
Speech: evoked potentials						
Phonological fusion						
Dichotic rivalry						
Ear advantage for speech						
Perception--auditory						
Language learning						
Action and perception						
Memory--digit span						
Individual differences						
Korean--stop consonants						
Voice Pitch and EMG						